

A Micro-Analysis of Seriation Skills

Denise M. Neapolitan

Ph.D.

**University of Edinburgh
1991**



Declaration

I declare that this thesis has been composed by myself and that the research reported therein has been conducted by myself unless otherwise indicated.

Denise M. Neapolitan

Edinburgh, 6th August, 1991

Acknowledgements

I would like to thank my supervisor Dr. Brendan McGonigle for his enthusiastic support and assistance throughout.

Also thanks to all at the Lab: Jim, Carlo, and especially Dr. Margaret Chalmers for her optimism in times of adversity. Thanks are due to the staff at the Edinburgh University Psychology Department Nursery, Sciennes Primary School, and Bruntsfield Primary School, and of course to all the children, for their participation. Dr. Michael Harris wrote the software used in the touch screen studies.

Special thanks are due to Randall Stark who donated his Macintosh SE/30 and all of his patience to the "cause." And finally, utmost thanks are due to my mother, without whose unconditional support this would not have been possible.

This thesis was supported by grants from the Marshall Commemoration Commission and the Alice Freeman Palmer Fellowship.

Abstract

The aim of this thesis is to make a contribution toward an eventual characterisation of causal factors in human cognitive development. This is accomplished by providing a rich, microanalytic description of change in a particular skill: seriation. The problem was tackled in two experimental phases, an extended clinical assay and a touch screen based assay. In the extended clinical assay, children aged 3 to 6 years were given the opportunity to review the task, locate errors, and repair them. The main finding was that intervention, for the most part, did not change performance --- subjects under age 4 years were entirely resistant to success, although some older subjects (aged 5 to 6 years) were able to succeed with assistance. In the touch screen based assay, subjects were given intensive training on variations of the seriation task suggested by McGonigle and Chalmers' decomposition of the seriation task. The scope of the thesis was limited to the serial control component of seriation (using arbitrary and non arbitrary strings). Subjects under age 4 years were focussed on for intensive training. Despite intensive training under immediate feedback conditions, it was found that subjects continued to be resistant to success. The source of subjects' limitations in reporting items, whether ordered arbitrarily or otherwise, was a limitation of serial control. Without the relevant underlying competence in place, training therefore appears to have only limited effect. That some older subjects can improve their performance indicates that only when the relevant competence is already in place can instruction be effective.

Table of Contents

Chapter 1: Introduction	1
<i>General Aims.....</i>	<i>1</i>
<i>Approaches to developmental change.....</i>	<i>1</i>
Piaget.....	1
Vygotsky.....	3
Uncovering early competence	4
Change restored.....	5
Conclusions.....	7
<i>Methodological context</i>	<i>8</i>
<i>Task domain: Seriation</i>	<i>10</i>
Criteria for selection.....	10
The task.....	10
Indexicality	10
Transparency.....	12
<i>The Experimental Programme</i>	<i>21</i>
Chapter 2: The extended clinical investigation.....	23
<i>Rationale.....</i>	<i>23</i>
<i>EXPERIMENT 1</i>	<i>25</i>
Subjects.....	25
Apparatus & Materials.....	25
Design.....	25
Procedures	26
<i>Results.....</i>	<i>30</i>
Comparison with Inhelder & Piaget.....	30
Breakdown of Performance.....	31
Age Distribution.....	35
Comparison of subject status pre- and post- intervention.....	35
Intervention Error Analysis.....	36
<i>Results Summary</i>	<i>40</i>
<i>Discussion.....</i>	<i>40</i>

**Chapter 3: New paradigms and procedures for the study of
seriation44**

Communicating task and error information44

Decomposing seriation skills46

New paradigms and procedures47

Chapter 4: Assessing serial ordering skills using the touch screen.....51

Rationale.....51

EXPERIMENT 2.....53

 Subjects.....53

 Apparatus & Stimuli.....53

 Design.....54

 Procedures55

Results.....59

 Acquisition.....59

 Verbal Probes62

 Over-training (Routinization).....64

 Test Phase I.....67

 Test Phase II.....74

Results Summary81

Discussion.....82

 Training82

 Verbal Protocols83

 Reaction Time in Over-training.....83

 Test Phase I.....84

 Test Phase II.....86

General Discussion.....88

**Chapter 5: Training and resistance points in robustly failing
subjects95**

Rationale.....95

EXPERIMENT 3.1.....97

 Subjects.....97

 Apparatus97

 Stimuli98

 Design.....98

 Procedures99

<i>Results</i>	102
Monotonic Size Series Acquisition.....	102
Verbal Probes	105
Over-training and Stimulus Size Set Test.....	106
<i>Results Summary</i>	109
<i>Discussion of Experiment 3.1</i>	110
EXPERIMENT 3.2	111
Subjects.....	111
Apparatus & Stimuli.....	111
Design & Procedures	112
<i>Results</i>	113
Arbitrary Colour Series Acquisition.....	113
Verbal Probes	117
Over-training.....	118
Comparison of Size Series and Colour Series Results.....	121
<i>Results Summary</i>	123
<i>Discussion of Experiment 3.2</i>	124
Chapter 6: Serial control and dimensional control in seriation tasks	127
<i>Rationale</i>	127
EXPERIMENT 4.1	129
Subjects.....	129
Apparatus & Stimuli.....	130
Design & Procedure.....	130
<i>Results</i>	131
Performance after the first two sessions	131
Performance after the second two sessions.....	133
Error analysis for Gill and Zach.....	133
RT analysis for Alexa and Yvonne	133
<i>Discussion of Experiment 4.1</i>	135
EXPERIMENT 4.2	136
Subjects.....	136
Apparatus & Stimuli.....	136
Design & Procedure.....	136
<i>Results (Andrew)</i>	140
Reverse Size Series Acquisition	140
Verbal Probe.....	141

Over-training.....	141
Trial Blocks Test.....	142
Reverse Colour Series Acquisition.....	144
Verbal Probe.....	144
Over-Training.....	145
Trial Blocks Test.....	146
<i>Results (Alexa & Yvonne)</i>	<i>147</i>
Refreshment.....	147
Acquisition of Reverse Series	147
Trial Blocks Test.....	148
<i>Results Summary.....</i>	<i>150</i>
<i>Discussion of Experiment 4.2</i>	<i>151</i>
<i>Summary of Experiments 3 & 4.....</i>	<i>153</i>
Chapter 7: Conclusions and issues for further research.....	155
<i>Conclusions about seriation.....</i>	<i>156</i>
Genuine competence limitations.....	156
Possible proximal zone.....	158
Resistance points.....	158
<i>Comparison to the Piagetian account of seriation.....</i>	<i>161</i>
<i>Conclusions about developmental change.....</i>	<i>163</i>
<i>Issues for further research.....</i>	<i>165</i>
The shift from blocks to the touch screen	165
Further exploration of young subjects' limitations.....	166
Exploration of subjects on the brink of success	167
<i>Closing.....</i>	<i>168</i>
References.....	170
Appendix A: State Evaluation Study Procedures and Results.....	I
Appendix B: Seriation with Reduced Sets.....	VII
Appendix C: Limited Review of Monotonic and Non-monotonic Series.....	X

List of Tables

Table 1.1: Results from Inhelder & Piaget, 1964, p. 250.....	11
Table 2.1: Comparison with Inhelder & Piaget.....	30
Table 2.2: Results after intervention compared with previous results	36
Table 2.3: Error Type.....	37
Table 2.4: Depth of Intervention.....	38
Table 4.1: Subjects' condition and string assignment.....	55
Table 4.2: Trials to Criterion: Training.....	60
Table 4.3: Trials to Criterion: Distribution.....	61
Table 4.4: Verbal Probes	63
Table 4.5: Average Inter-item RT expressed in hundredths of seconds (percent of decision time in parenthesis).....	65
Table 4.6: Test Performance by Subject.....	68
Table 4.7: Performance on quadruplets, triplets, and pairs.....	69
Table 4.8: Breakdown of performance by "gaps"	70
Table 4.9 : Breakdown of performance on A-start, B-start, and C-start trials	70
Table 4.10: Comparison to Terrace and D'Amato & Colombo.....	72
Table 4.11: Comparison to Terrace and D'Amato & Colombo (common trials only)	74
Table 4.12: Test Phase 2 Summary.....	76
Table 4.13: Error Type Summary	80
Table 5.1: Subjects' age and string assignment.....	99
Table 5.2: Trials to criterion: size series training.....	103
Table 5.3: Distribution of trials to criterion: size series training.....	104

Table 5.4: Size series verbal probes.....	106
Table 5.5: Mean decision time and IRT expressed in hundredths of seconds (% of decision time in parenthesis).....	108
Table 5.6: Trials to criterion: colour series training.....	113
Table 5.7: Distribution of trials to criterion: colour series training.....	114
Table 5.8: Comparison with subjects in Experiment 2: trials to criterion.....	115
Table 5.9: Comparison with subjects in Experiment 2: distribution of trials to criterion.....	116
Table 5.10: Colour series verbal probes.....	118
Table 5.11: Mean decision time and IRT expressed in hundredths of seconds (% of decision time in parenthesis).....	119
Table 5.12: Comparison of size and colour series.....	122
Table 6.1: Trial Blocks Test performance	132
Table 6.2: Mean decision time and IRT (% of decision time in parenthesis).....	134
Table 6.3: Trial blocks test performance, size series (Andrew).....	143
Table 6.4: Trial blocks test performance, colour series (Andrew).....	146
Table 6.5: Experiment 4.2 performance	148
Table 6.6: Mean Decision Time and Inter-Item RT (% of Decision Time in parenthesis).....	149

List of Figures

Figure 1.1: The ordinal rule learning paradigm.....	17
Figure 1.2: The serial learning paradigm.....	17
Figure 1.3: The serial X ordinal (combined) learning paradigm (monotonic)	18
Figure 1.4: The serial X ordinal (combined) learning paradigm (non-monotonic)	19
Figure 2.1: Design.....	26
Figure 2.2: A subject performing the seriation task.....	27
Figure 2.3: Review Procedure.....	28
Figure 2.4: Intervention Procedure.....	29
Figure 2.5: Breakdown of performance by functional taxonomic group.....	33
Figure 2.6: Age distribution of functional taxonomic groups.....	34
Figure 3.1: The paradigms.....	48
Figure 3.2: A subject performing the serial order task on the touch screen.....	49
Figure 4.1: Inter-item RT as a percent of decision time: Forward v. Backward training.....	67
Figure 4.2: Series Completion: Distribution of total error	77
Figure 4.3: Series Completion v. Next item correct.....	78
Figure 5.1: Mean IRT on training set trials as a % of decision time (all subjects except Gill).....	109
Figure 5.2: Mean IRT as a % of decision time (group).	120
Figure 5.3: Inter-item RT as a % of decision time (individual subjects).....	121
Figure 6.1: Experiment 4.1 Design	131

Figure 6.2: Size v. colour series IRT (Alexa).....134

Figure 6.3: Size v. colour series IRT (Yvonne)135

Figure 6.4: Experiment 4.2 design (Andrew).....138

Figure 6.5: Experiment 4.2 design (Alexa & Yvonne)140

Figure 6.6: Comparison of IRT for 2 size series (Andrew).142

Figure 6.7: Comparison of IRT for 2 colour series (Andrew).146

Chapter 1: Introduction

General Aims

Human beings manage to achieve amazing cognitive feats over the long period of time which spans virtually helpless infancy to fully functioning adulthood. But what are the causal factors which determine cognitive change throughout the long trajectory of development? The aim of this thesis is to make a contribution toward to this problem. This is accomplished by looking intensively at change in a particular skill: seriation. The goal of this chapter is to introduce and support the experimental programme of the thesis. First however, the issue of developmental change and how it has been investigated must be addressed. There is considerable debate over the issue of *what* changes in development. Secondly, how this debate has influenced the methodology used in this thesis is discussed. Thirdly, the choice of seriation as a task domain is discussed. Finally, the experimental programme of the thesis is presented.

Approaches to developmental change

Piaget

Jean Piaget and his co-workers have had a great influence the study of cognitive development (see Piaget, 1970; 1974 for an overview). Piaget's theory of development is a *stage theory*. Development is achieved by successive stages. There are four major stages of cognitive development. The first stage is that of sensorimotor intelligence lasting from birth until about age eighteen months. The second stage is that of preoperational representation which begins at age eighteen months and lasts until age seven or eight years. Between ages seven and twelve years is the stage of concrete operations. The final stage beginning after age twelve years is that of formal operations. The stages reflect different levels of competence. Each stage is built upon the structures of the one before it. Thus development may be seen as a series of steps with plateaus. Piaget saw cognitive development as "epigenetic." Changes in competence are determined by the child's action and interaction with the environment.

Through the processes of assimilation and accommodation, actions become interiorized. These interiorized actions form operations, which are the basic structures of thought.

Piaget built his theory upon a large range of observations of children's cognitive abilities. Because Piaget saw the end-state of development being thought guided by logical operations, he devised a series of tasks designed to diagnose the child predecessors of logical operations. Particular types of performance served to place a child at a particular stage in development. The inference to be drawn in the case of default was that the child was restricted in competence. Piaget's investigative technique is known as the *clinical method*, which typically involved asking the child questions in the context of a single experimental setting. In the 1960's researchers began to conduct investigations attempting to replicate Piaget's results using standard experimental procedures (see Flavell, 1963 for a review). These confirmed the robustness of the Piagetian results in many content areas including number (Dodwell, 1960; 1961), conservation (Elkind, 1961a; 1961b), and seriation (Elkind, 1964).

Despite the robustness of the results, Piaget failed to provide an adequate characterisation of the transition from stage to stage. Piaget outlined the mechanism of transition from stage to stage as one of adaptation to the environment by means of assimilation and accommodation. However, this has proven to be unsatisfactory. As Flavell (1963) has argued, Piaget has "shed little empirical hard-fact light on precisely how these forms [cognitive structures] work their way into the child's cognitive life" (p. 370). Piaget, being interested in constructing a global theory of cognitive development, concentrated on cross-sectional studies revealing macro-change in age groupings of children rather than a micro-analysis of individuals. Further, the role of experience outwith the "lone discovery" of the child was left largely unexplored by Piaget. The stage structure limited the possibility of improving performance by instruction because an implication of the successive nature of stages is that children cannot assimilate or accommodate that which is incompatible with their present system of understanding. Thus, instruction could at best produce limited or temporary advances which had no generality beyond the area being trained.

Vygotsky

L. S. Vygotsky (1934/1962; 1978) defined a different kind of stage theory, which is in many respects similar to Piaget's. Development is characterised by revolutionary shifts in competence. The most important is the move from "elementary" mental function to "higher" mental functions. Major transition points are defined in terms of the form of mediation utilized. Most important for Vygotsky is the mediation of thought by tools and signs, with language being the most important of these. The main difference between Vygotsky and Piaget relevant here is that Vygotsky saw the child's interaction with others, rather than the child's self-motivated interaction with the environment, as a key determinant of change.

Vygotsky's theory of cognitive development rests heavily on the notion of *internalization*. For Vygotsky "all higher mental functions are internalized social relationships," (1981, p. 164, cited in Wertsch, 1985a, p. 66). Of particular importance in cognitive development are the relationships between children and adults. First the adult directs the child's activity. Gradually the child takes the initiative and begins taking control of the activity with adult correction and guidance when required. Finally the child acts independently.

Vygotsky introduced the notion of the *Zone of Proximal Development* (1978, Chapter 6) to describe the gradual transition from inter- to intrapsychological functioning. Although adult assistance is matched in some sense to the child's present developmental level, the child can be helped to progress beyond that level. In order to provide assistance, one must not only know the current intellectual level the child has attained, but also the child's potential level of performance when given guidance to help with the problem at hand. The Zone of Proximal Development is the difference between these two levels, "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (1978, p. 86).

In this scenario, task default must be interpreted in two contexts: independent problem solving and assisted problem solving. If default occurs during independent problem solving, it is unclear whether the child is restricted in competence. It is possible that the child may be able to achieve success through social interaction. However if default occurs during assisted problem solving, the inference is that the child is restricted in competence. A number of recent studies have used the notion of the Zone of Proximal Development to help diagnose and explore possible transition points in development (see Wertsch, 1985b). However these studies have been restricted to particular domains (e.g. reading) and have not given rise to global characterisations of development.

Uncovering early competence

Taking a lead from Chomsky's ideas (1965) about the innateness of language structures, one focus of contemporary research has been to reject the idea of competence changes in development. In this scenario, development essentially involves children's increased access to innately given competences. The predecessors of adult thought are present in child thinking but are restricted in application. Young children's deficiencies are not the result of a lack of competence, but the failure of performance (Gelman & Gallistel, 1978).

The general methodology has been to modify Piaget's tasks to reveal early competences. The landmark study here is McGarrigle & Donaldson's (1975) "naughty teddy" experiment. Four and five year old children's judgements on conservation problems were studied under modified conditions of the standard Piagetian tasks. In this case the transformation of the quantity was not performed deliberately by the experimenter but accidentally by a "naughty" teddy bear. Under these conditions it was found that 63% of children could conserve as opposed to 16% under the standard conditions. These results were interpreted as indicating that "traditional procedures for assessing conservation seriously underestimate the child's knowledge," (1975, p. 347). Early competences have been sought through the modification of other task domains such as classification, number concepts, arithmetic and measurement (see reviews by Gelman, 1978; Donaldson, Grieve, & Pratt, 1983).

Donaldson (1978) argued that children's true competences can only be revealed in situations which make *human sense* to the child, that is when the task is put in a context which is fully intelligible to the child. In this sense the failure in the standard tasks is not the failure of the child, but the failure of the experimenter to create a task environment in which children's underlying competences can be uncovered.

There are a number of problems with this approach. Light (1986) has pointed out a number of methodological problems with modified conservation tasks. With respect to context, "incidental" or "accidental" transformations may be ignored by the child (the "chipped beaker" study, Light et al., 1979). Also Light and Gilmore (1983) found that contextual modifications which lead children to "success" can also lead them into error in situations where the quantity in question (area) is not in fact conserved. There is a deeper problem however with the idea of fixed competence. The argument that the task was not appropriately contextualized to reveal the child's underlying competences can always be made. Child deficiencies can always be cited as performance limitations. This renders developmental phenomena virtually untestable. As a consequence, such studies offer no effective prescriptions for how to specify the course of development or what is meant by "increasing access" to competence. In this sense, real developmental change has been "argued away" by contemporary research.

Change restored

Recent evidence, both from behavioural and neuropsychological quarters, indicates that developmental researchers should take seriously the idea of real changes in competence in development. Recently Sugarman (1987) has called for a new agenda for developmental research. She argues that the most basic goal of developmental psychology should be "to describe the child's mind and how it changes." The problem with existing work, she argues, is that developmental research has been over-concerned with the endstate. Development has been seen as a process in which the simple forms of child thought are transformed (or extended) in a continuous line into corresponding adult forms. This conception, implying the need to causally link early behaviours with later behaviours, has obscured

research into the essentials of child thought and how it develops into adult thought.

In order to rectify the situation, Sugarman argues that serious descriptive work needs to be done and offers guidelines for further research. In particular, she argues that researchers "treat development as a first assumption and not as a postulate to be proved or disproved" (p. 23). Proving or disproving whether or not there are real changes in competence is virtually intractable. Studies should take as a working assumption that real change occurs in development in order that novel hypotheses can be formulated about how children think. In this respect, developmental change data should be actively sought by researchers. Thus, an important aim in research is "to find tasks task that expose nontrivial differences in the performance of older and younger children," (p. 34). Further, Sugarman argues that research must treat the endstate of development as the result and not the premise of the investigation. In other words, children's behaviour should be described and assessed for what it is, not for how it deviates from the endstate. The benefit in doing this is that children's thought will be more accurately described. Also, it may be possible to discover properties of the endstate that have not previously been identified.

Other recent research also supports a view of development in which there are age-related changes in competence which are not necessarily causally linked. In a review of studies of continuities and discontinuities in development, Emde and Harmon (1984) conclude that there is a trend toward characterising development as multidimensional and non-linear, "early behavioural change processes occur which may or may not be connected with later life behavioural change processes," (p. 2). Inferences about the biological bases of change are also emphasized. In the same volume, Kagan (1984) infers from research on the emergence of memory and self-awareness that "major cognitive competences emerge as a consequence of maturation of the central nervous system, rather than gradually acquired through a long series of interactions," (p. 28).

Although there is behavioural support for distinct stages of human development, there has been little neurophysiological evidence of discontinuities and growth spurts during childhood and adolescence.

However, recent advances in neuropsychological research have provided a neural basis for major spurts and qualitative shifts in development. Electrophysiological studies have provided some evidence for developmental staging of the human cerebral hemispheres. Thatcher, Walker, & Guidice (1987) have recently assessed the development of the cerebral hemispheres using measures of EEG coherence and phase in 577 children ranging in age from two months to early adulthood. They found discrete growth spurts that appeared in specific anatomical locations at specific periods in development. Their results showed increase in phase from birth to about age three years for both right and left hemispheres. From age four to six years the left hemisphere showed a marked increase in phase that tended to level off from age six years to adulthood. A slight growth spurt in the right hemisphere was found between the ages of eight and ten years. Also there was a weak (not statistically significant) bilateral growth spurt around ages eleven to thirteen years and from sixteen years to adulthood. Thatcher et al. interpret their results as indicating "a sequencing of different anatomical systems during postnatal cortical development," (p. 1113). Furthermore the timing of EEG changes overlap with the timing of the Piaget stages. These results support an ontogenetic view of cognitive development, that is, by the "genetically programmed unfolding of specific brain functions and specific brain connections," (p. 1110).

Conclusions

What can be concluded from this review of the literature? Theories which call for qualitative change in development make sense from both a behavioural and a biological point of view. Sugarman has argued that there are sound behavioural reasons to support the idea of competence changes in development. Further, researchers in the neurosciences, in demonstrating spurts in brain growth which correspond with the ages specified by the Piagetian stage theory, have shown that there are neuropsychological reasons to support the idea of competence changes in development. An implication of these results is that early behaviour is not necessarily antecedent, that is, causally linked to and continuous with later behaviour (McGonigle & Chalmers, in press). This relieves the need to explain early behaviour (on a particular task) in terms of its contribution to later behaviour. New brain growth means new elements

(competences) may be entering the cognitive system which are not necessarily causally linked to what was there before. Thus development may reasonably be seen as a progression of qualitative changes in competence which may be discontinuous with previous levels of competence.

Methodological context

In order to provide support for or against such a view of development it is necessary to examine behaviour which is *indexical* of cognitive change in such a way as to make *transparent* to the experimenter the causal factors underwriting change.

Because it would be impractical to study all of development, it is necessary to locate particular task domains which can "stand proxy" for development. This means that task domains must be sensitive to cognitive change throughout the developmental trajectory, as determined by behavioural and neuropsychological change. Further if a task domain is to serve as a "window into the mind" it should reflect important cognitive processes. If the chosen domain is an isolated or trivial process, however detailed and sophisticated the analysis, it will provide a "window" which is too small to provide useful information concerning its general structure. Finally, behaviour elicited by tasks should be reliable, that is capable of replication.

Piaget located task domains which were at least putatively reliable and indexical of changes in competence (see Flavell, 1963 for a review of replication studies). However, he failed to make them sufficiently transparent. What is needed is procedures which get beyond description in order to discover causal factors. Piaget focussed on the macro-analysis of children's behaviour. His tests were diagnostic with respect to the stage theory. Particular types of default located a child's place with respect to a particular stage or substage of development. Description and taxonomy of behaviour were motivated by his characterisation of the end-state of development as thought guided by logical operations. Because Piaget was so focussed on this characterisation, he tended to "impoverish" descriptions of what individual children were doing in order that the behaviour should conform to a global theory of child development (Sugarman, 1987).

Piaget's protocols of children's performance (within stages) lack important information. Piaget's results were largely cross-sectional. Tasks were administered in a single experimental session. There was no follow up on the child, nor did he trace individual children through the stages. Subsequent research has also tended to concentrate on cross-sectional studies. These yield grouped and averaged data. These results provide useful "snapshots" of behaviour, however characterisation of transitions in competence is difficult. Grouped and averaged data lack the kind of detailed information needed to outline what is happening during periods of transition. Further knowledge about individual children's abilities especially with respect to error might also lead to insights about the sources of default. This requires a detailed investigation of the types of errors subjects make at different points in development and how they use error information. What is required is a detailed micro-analysis of the behaviour of individuals: not snapshots of individuals, but a behaviour graph of their performance.

Returning to the idea of a neurological connection in cognitive development (Thatcher et al, 1987) it is necessary to find a way to map new elements of brain growth onto changes in cognitive competence. This problem can be attacked by the analysis of cognitive skills into functionally separate and independent components. It is then necessary to discover what cognitive resources children start with, which are added along the developmental trajectory and how these components interact one with the other (McGonigle & Chalmers, in press). By this characterisation, success at a skill is due to the synthesis of cognitive subsystems which are integrated through learning and development, and failure (or default) at a skill is due to the absence or non-integration of one or several components. An implication is that errors are not all made for the same reasons: there is no one critical rule whose absence will cause default. Indeed, even with all of the components present failure may still result due to their non-integration. Of course, key to this whole enterprise is discovering what the components are and how they assemble themselves into a system.

Task domain: Seriation

Criteria for selection

What tasks are good candidates for micro-analysis? McGonigle and Chalmers (in press) have suggested that candidate task domains should be evaluated using two criteria: indexicality and transparency.

Indexicality. Candidate task domains should be putatively indexical of changes in cognitive competence. Such changes should span substantial periods of growth (e.g. with respect to the time scale for brain growth outlined by Thatcher et al., 1987). If a task domain is to "stand proxy" for development, candidate task domains must exhibit changes in behaviour over a period of time largely coextensive with the time scale of development and should tap into cognitive processes which are ubiquitous, important, and typical.

Transparency. Once a task or class of tasks has been identified as indexical of changes in competence, it is necessary to ensure that the behaviour subjects exhibit when engaging the task reveals these changes in a transparent fashion. In particular, solutions should permit the investigator to make inferences about component subskills of the task and their inter-relations with one another. Although the components of a domain must be accessible to the investigator, they need not be available to the subject for conscious access or verbal restatement.

The task

The task domain used in this thesis is *seriation*. The seriation class of tasks, used by Piaget and others, is an important indicator of cognitive growth. Typically the child is confronted with a monotonically ordered series of (usually ten) blocks or rods varying in size. The subject is required to construct a copy from a jumbled array. Ideal performance is a serial production in which the subject takes the biggest (or smallest) item first, the next biggest items second, etc., in one smooth serial production.

Indexicality

Best known are the results of Inhelder and Piaget (1964).

Age	4	5	6	7	8
no. of subjects	15	34	32	32	21
1A. No attempt at seriation (%)	53	18	7	0	0
1B. Small unco-ordinated series (%)	47	61	34	22	0
2. Success by trial & error (%)	0	12	25	15	5
3. Success with operational method (%)	0	9	34	63	95

Table 1.1: Results from Inhelder & Piaget, 1964, p. 250

The results show change in seriation behaviour taking place over a long time scale up to age eight years, a substantial portion of development. They divide seriation behaviour into four categories. There are two success modes and two failure modes. These are associated with age. Typically children aged four to five years cannot seriate at all or construct only small uncoordinated sets. Children aged five to six years achieve solution via trial and error. Finally, children aged seven to eight years seriate successfully in one principled production --- arranging all items in rank order in one "go."

The seriation task domain was important for Piaget because the task of assembling an ordered series of blocks of (e.g.) monotonically increasing sizes served as a concrete behavioural analog of the logical operation of seriation. The operation of seriation is defined as the assembly of elements into a transitive, asymmetrical series of the form $A > B$, $B > C$, $C > D$, etc., and is a core operation of Grouping V, one of the nine groupings of logical operations which characterise operational thought (Flavell, 1963). Piaget interpreted variations in behaviour on the concrete task of seriation as indexical of changes in operational competence.

The preoperational child (seriation stage 1A) fails to correctly complete the series. When trying to find the position of an item, children must compare the item with those which have already been placed and those which are awaiting placement. Subjects at this stage fail to coordinate

successive asymmetrical relations into a "whole," that is that $A > B$, $B > C$ is equivalent to $A > B > C$ from which can be inferred the superordinate (transitive relation) $A > C$. The central difficulty is the child's failure to grasp the reversibility inherent in systems of asymmetrical relations, that is that each item in the series is (e.g.) both bigger than the item before it and smaller than the item after it. Some children (seriation stage 1B) are able to make isolated comparisons, but often fail to maintain a stable direction in the relationship between the items. This results in small series of two or three items displayed in a line.

Later in the preoperational stage subjects may achieve a correct solution by trial and error (seriation stage 2). The child can anticipate the global result, but not the steps which are necessary to obtain it. Although the child may end up with a correct copy of the model, this is achieved through use of intuition instead of logical order, i.e. perceptual comparison instead of logical operations.

The concrete operational child can correctly and systematically construct a series (seriation stage 3). Children are aware of the reversibility of relations inherent in the series. They use this principle to smoothly and simultaneously coordinate each successive relational comparison into a seriated whole. Furthermore this allows them to be able to insert new elements into a preformed series, without trial and error. The formal operational child can perform "mental" seriations on non-concrete, hypothetical sets, as in verbal transitivity problems (e.g., the Edith, Susanne, and Lilly problem).

The results obtained by Inhelder & Piaget have been shown to be reliable. The Piagetian seriation experiments have been replicated using standard experimental techniques (Elkind, 1964; Young, 1976, 1978) and the results of Inhelder & Piaget were confirmed. There is no study known at the time of writing which contradicts the Piagetian results using the standard ten rods as stimuli. Thus, at the macro-level seriation appears to be indexical of changes in competence.

Transparency

Having established that seriation behaviour is putatively indexical, the next question is: can seriation behaviour be made transparent? As

usually administered the seriation task involves subjects' manipulation of blocks. This means that seriation is a public act and solutions are visible to the experimenter. This provides an opportunity to "get a hook into" components of the task. However, analysis of seriation behaviour thus far has failed to make the task transparent enough to provide the investigator with insights into the transition from children's early inadequate behaviour to later principled success.

Piaget's analysis of seriation rests on the acquisition of a key rule, reversibility, which is essentially a relation among relations. This allows the subject to simultaneously coordinate all the relations and achieve operational success. Failure at any stage signifies a failure to coordinate all the relations and hence an absence of the key rule, reversibility. However in the early stages of seriation subjects can coordinate a few relations, resulting in "partial seriation" and uncoordinated series. Later, as the ability to coordinate more relations develops through the child's exploration and action, the child can achieve correct, but not operational, solutions by trial and error (1964, p. 258). Finally the simultaneous (two-way) coordination of all the relations develops. This, in Piaget's eyes, is true success; the culmination of previous activity by children and their resultant understanding of relations.

What this boils down to is that through action and discovery the child is able to coordinate increasingly more relations and hence increasingly more blocks. The more relations children can coordinate, the more blocks they can control in a series. This amounts to a tautology between behaviour and analysis --- fewer blocks, fewer relations, more blocks, more relations --- and hence, an inadequate characterisation of transitions. Because Piaget was keen to show how early activity and early structures culminated in the end-result, he did not pay enough attention to the particulars of children's limitations and abilities in the "early" stages of seriation, other than with respect to relational coordination. Exploration of these limitations and abilities might serve to make seriation more transparent.

Richard Young (1976, 1978) sought to solve the problem of transitions between stages of seriation behaviour by examining in detail the seriation behaviour of children at various points along the developmental

progression, and attempting to create explicit models of the mechanisms responsible for their performance. He hypothesised that the child's behaviour on a seriation task is determined by three aspects of the skill --- selection, evaluation, and placement --- and pays particular attention to these. Young proposed to "extend" the Newell & Simon (1972) notion of a protocol as "think aloud" data (subjects describing their path through the problem space) to the frame-by-frame analysis of the child's every move on video tape (Young, 1978, p. 361). He believed that this method would allow him to infer the strategies which underwrite children's seriation behaviour. Based on these observations Young wrote a production system to model the child's behaviour.

However Young never outlined a proper task analysis of seriation. He concentrated on describing the surface features of seriation performance and never really questioned what skills might underwrite the selection of blocks or their correct placement. This resulted in a description of seriation which was more detailed but no richer than Piaget's. "The picture of seriation development suggested by these analyses is that the child begins with just the ability to arrange blocks in a line, and then gradually acquires the rules that lead him through one or more of the observed pre-seriation phenomena, on to simple seriation, and finally more reliable seriation," (p. 210-11). Young's model of seriation (a system of production rules) relied on the addition of critical rules, similar to Piaget's notion of reversibility as critical to operational seriation. "[...] differences in PSs [production systems] have been due to the inclusion of additional rules. In some cases the difference between being able or unable to seriate lay in the addition of just one critical rule," (p. 209). Young also allowed the end-state of seriation success to intrude upon his description of younger children. He described what young subjects were doing in terms of its role in successful seriation. "The various phenomena observed in non-seriators -- unordered lines, subseries, partial seriations, and so -- are not due to special pre-seriation processes, but instead arise predictably from PSs which differ from those of successful seriators only in lacking one or more rules," (p. 208-9). However Young does not address the issue of where rules come from and how they get added to the production system.

A particular problem with the accounts both Piaget and Young is their description and categorization of default. The end-state of successful seriation impinges on this description. Take for example, the category of small uncoordinated series or partial series. Such series are "small," "uncoordinated," and partial only with respect to a successful seriation. This doesn't really describe what the child was doing. It is an interpretation of what the child may have been trying to do. In order to make seriation more transparent what is required are procedures which get beyond description of the surface. In particular not enough attention has been paid to teasing out the error factors determining default. Both Piaget and Young concentrated on critical rules, however a second look at seriation and the requirements of the task indicates that default is likely to be multifactorial.

What is required are tasks which identify error factors. Using an error motivated theory of description, it is possible to design procedures to partition the error space in order to achieve a more useful description of why children fail. This is the approach taken in this thesis.

The problem can be examined first from the point of view of the subject. Piaget assumed that the seriation task was universally understood because it was concrete and perceptible. Therefore he did not enquire into the child's understanding of the task. But it must be asked whether the subject is trying to solve the same task as the experimenter? Sonoda (unpublished manuscript) investigated the self-corrective abilities of young children on a seriation task and found that subjects' seriation ability corresponded to their ability to spontaneously monitor their own actions, specifically with respect to error location and repair. An implication is that the same error factors may not be at work throughout the developmental range. Children may fail for different reasons at different stages of development. Awareness of error may have an important role to play in change (see Karmiloff-Smith, 1986). Marshall & Morton (1978) in their model of speech hypothesize that, in development, awareness of error is "a crucial part of the (internal) education system," (p. 233). If subjects are not aware of their mistakes they have no behavioural motive to change them. The result is an incorrect production or "failure." However, if subjects are not self-motivated to change it may be possible to

induce change by providing feedback and assistance socially, as per Vygotsky's Zone of Proximal Development. In this scenario it is possible to investigate to what degree children are feedback informed, what error signals to them at different points in development, and what it prompts them to do (in terms of repair).

Next the error space can be examined from the point of view of the experimenter. In order to get a clear picture of the complex error space of seriation what is required is an analysis of the seriation task into component skills. Skill decomposition has several advantages. Firstly, with skill decomposition, it is possible to precisely describe what failure is signaling; what are the possible loci of default. Secondly, with an hypothesis about what skills make up seriation, it is possible to examine individual subjects longitudinally on variations of the seriation task suggested by a particular skill decomposition. In this way, it is possible to diagnose resistance points to success in the individual. Finally, it is possible to assess which component skills are capable of being learnt (and which are not) using standard learning procedures. In this way we might discover how and when, over the course of development, components are assembled into the top-level skill of seriation. McGonigle & Chalmers have initiated an experimental programme designed to diagnose faulty components of seriation. They have decomposed the seriation task into several components including serial order, ordinal computation, modeling (copying) and search (McGonigle & Chalmers, in press; McGonigle, 1987). This decompositional task analysis has inspired new paradigms in which to study these components first in isolation and then in coalition (see McGonigle, 1987). These are shown schematically in Figures 1.1, 1.2, 1.3, and 1.4.

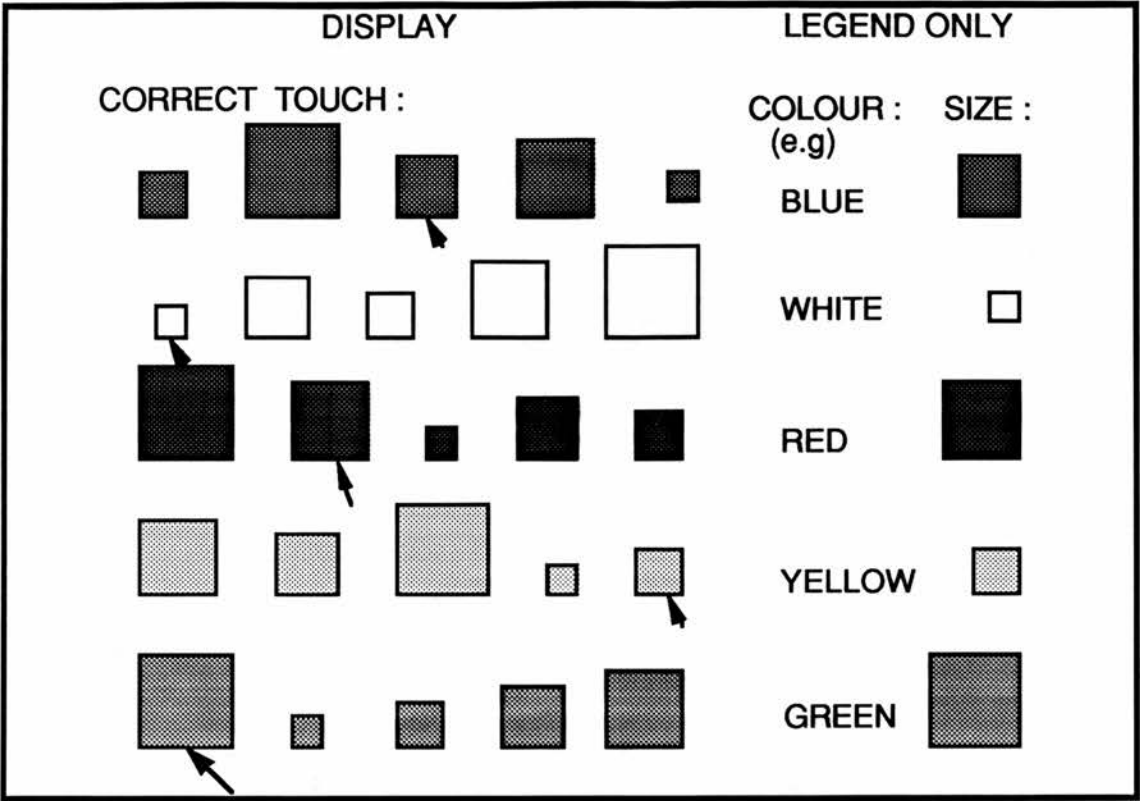


Figure 1.1: The ordinal rule learning paradigm

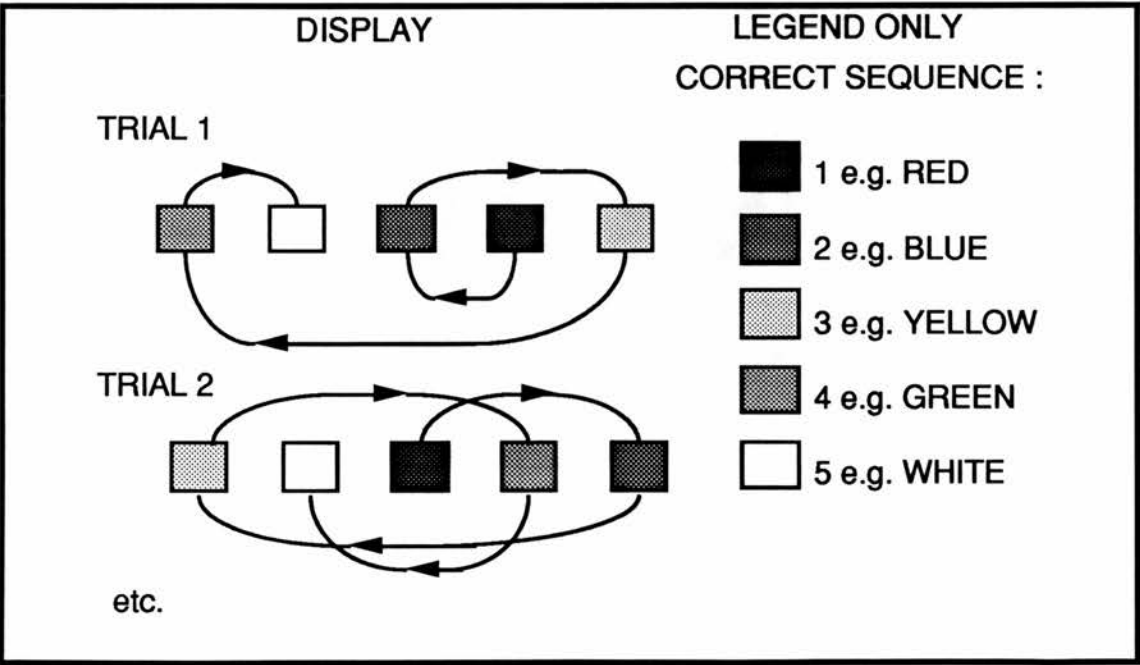


Figure 1.2: The serial learning paradigm

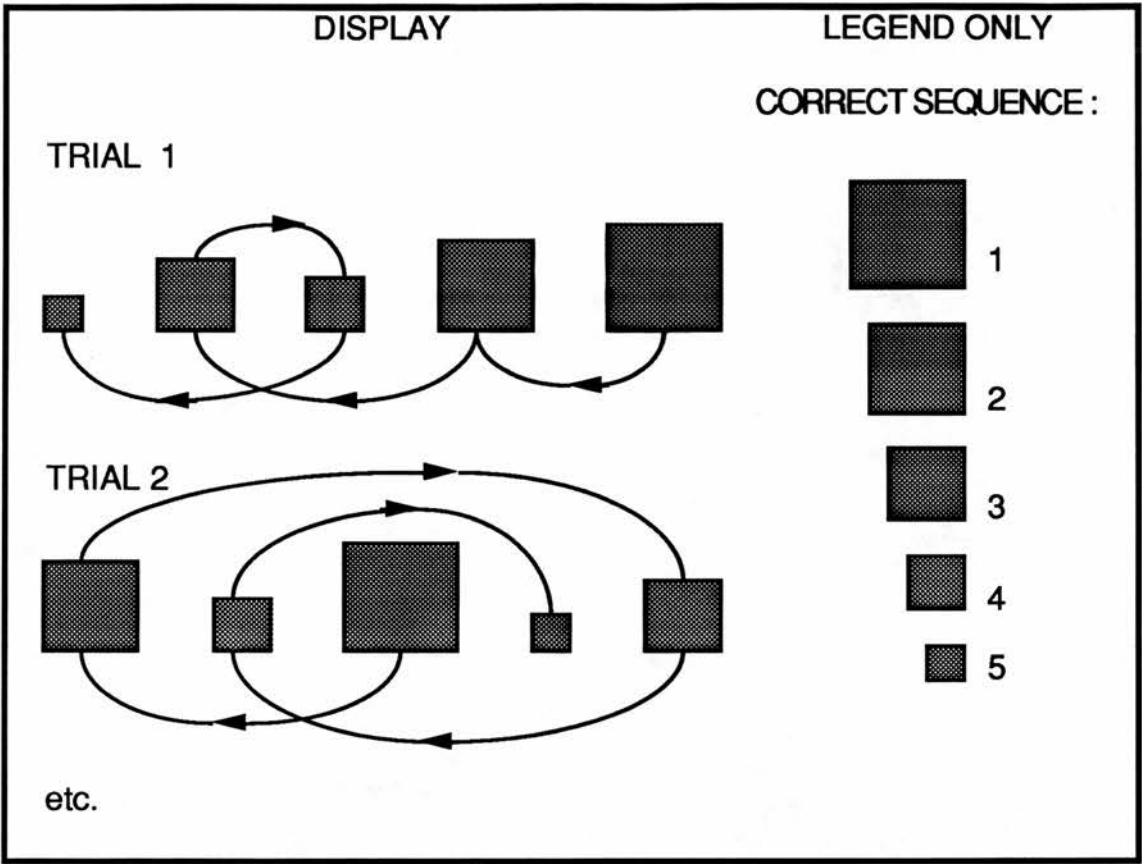


Figure 1.3: The serial X ordinal (combined) learning paradigm (monotonic)

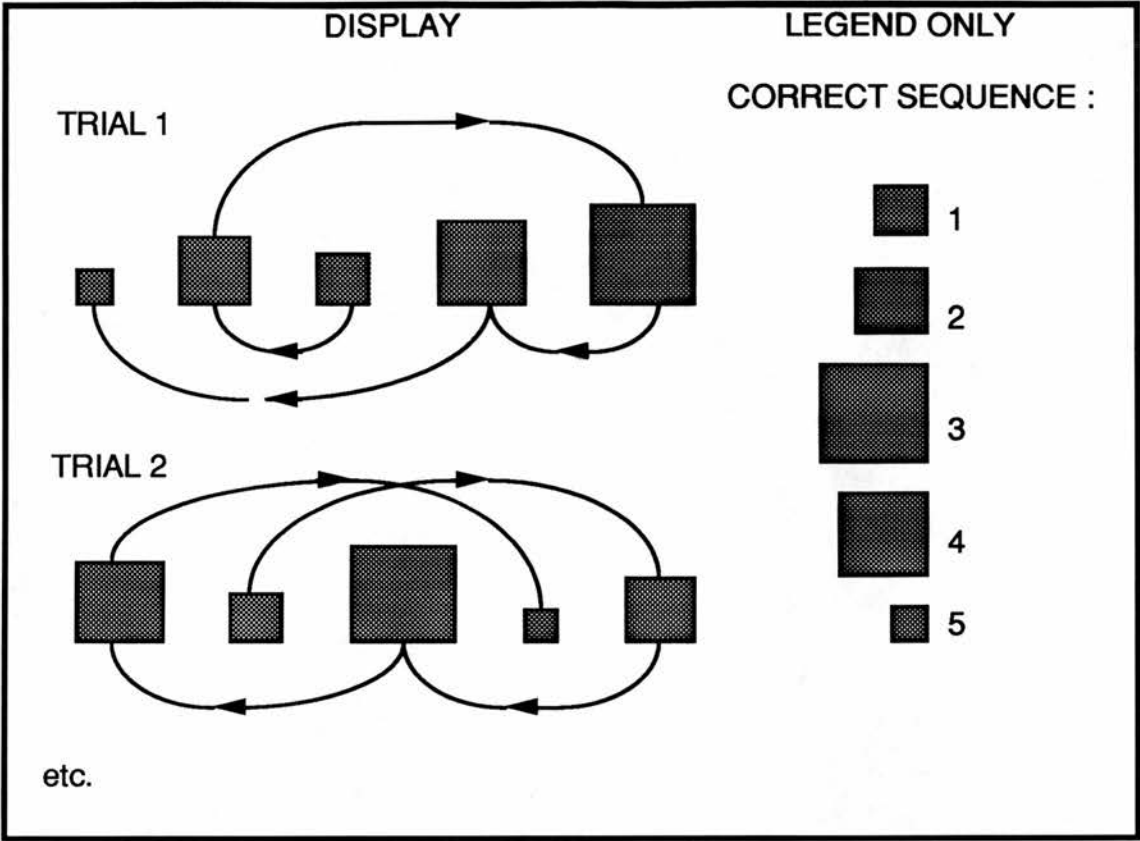


Figure 1.4: The serial X ordinal (combined) learning paradigm (non-monotonic)

The ordinal component has been isolated from the modeling and serial production aspects of the task (McGonigle & Chalmers, 1986). An ordinal rule task (see Figure 1.1) was devised to determine whether stimulus relations in the model can be perceived by subjects otherwise incapable of constructing the series (monkeys and young children). In this task the child is confronted with five blocks varying in size. The position and configuration of blocks changes from trial to trial. Each ordinal rule to be learnt is denoted by a different colour. For example, if all objects are red then select the biggest block. The subject discovers this rule by trial and error: children found counters under the correct objects. When this rule is learnt to a performance criterion, a second rule is taught, etc., until all rules are in the subject's repertoire. At this point the subject is tested on all rules given together in a random sequence.

McGonigle and Chalmers found that ordinal computation is possible for young children who cannot model a series. This raises questions about how young children would perform on tasks designed to separately assay the modeling and serial production aspects of seriation. The fact that pigeons, a species with minimal neocortical brain development, have been shown to manage information via serial order (Straub et al., 1979; Straub & Terrace, 1981), indicates that it is likely to be a "design primitive" or basic structure rather than a higher order benchmark of cognitive achievement (McGonigle, 1987). However there is also the question of how the components interact as a system. In seriation it is necessary to use each ordinal rule in the set in a particular serial order, e.g. from biggest to smallest. However, subjects may be able to identify the ordinal position of an object within a series without being able to use the rules together as a system. Having a repertoire of ordinal rules may not be a sufficient condition for seriation. An analogous situation can be found in early linguistic development where children may have sizable vocabularies but are still limited to very short "sentences." (McGonigle, 1987).

McGonigle and Chalmers have designed a training paradigm to investigate the interaction of ordinal and serial components (see Figures 1.3 and 1.4). This paradigm features a particular apparatus, a computer equipped with a touch-sensitive screen. On the screen are displayed a set of objects differing in size. Touching an object results in a registration

tone. An important feature of this paradigm is intensive training using parsed input and immediate feedback. The subject is required to discover (by touching) which sequence of objects is correct. For example, the subject may be required to compute the ordinal rules from biggest to smallest, in the monotonic case. Alternatively, subjects may be required to compute the ordinal rules in a non-monotonic order. Incorrect touches result in a negative feedback tone. A correctly completed sequence is signalled by a positive feedback tone. The position of objects changes from trial to trial. Using this paradigm it is possible to explore the interaction between serial and ordinal components of seriation. Also it is possible to investigate whether or not these skills are trainable in young children who spontaneously fail in classical versions of the seriation task.

The Experimental Programme

This scenario leads to an experimental programme which proceeds in two phases of research: the extended clinical assay and the touch screen based assay. The first phase of research involves the extension of classical clinical methods of testing children. The first goal of this phase is to check the reliability of Inhelder & Piaget's findings; that the pattern of age-wise stages still stands. The second is to assess the role of behavioural factors in change by providing simple intervention, e.g. calling the child's attention to error and giving them an opportunity to repair. The question of whether (and at which ages) children's failures are robust and real is addressed, as well as whether and when they can improve their performance through intervention. Finally, this work centers on the identification of reasons for failure. Are the same error factors at work throughout the developmental range, or do children fail for different reasons at different stages of development? The results of this phase are presented in Chapter 2.

The touch screen based assay is so called due to the apparatus used --- a computer equipped with a touch sensitive screen. If the extended clinical assay shows that failure is robust and real, we are entitled to enter the second phase of research using the McGonigle & Chalmers compositional task analysis of seriation and associated paradigms. This phase of research is focussed on task components assayed in a learning situation where error information is immediate. Our aim is to gather further information about the crucial loci of change, what precisely is

changing and how. In this thesis we limit ourselves to the serial production component of seriation. The main tension is between serial abilities alone using unrelated stimuli (coloured squares) and serial abilities combined with ordinally related stimuli (sized squares). The results of this phase are presented in Chapters 4, 5, and 6.

Thus the plan of the thesis is as follows. The results of the extended clinical investigation are presented in Chapter 2. Chapters 4, 5, and 6 are concerned with touch screen studies. Chapter 4 represents the first use of the touch screen and presents a study of serial ordering skills in four year olds. Chapter 5 presents an intensive study of the serial and seriation abilities of children younger than four years. Chapter 6 presents a study designed to address questions about the use of serial and dimensional control raised by the performance of some subjects in Chapter 5. Finally Chapter 7 discusses the picture of the development of seriation which emerges from these studies, and discusses implications for the larger issue of the nature of change in cognitive development.

Chapter 2: The extended clinical investigation

Rationale

Before entering into any extensive investigation, it is necessary to check Inhelder & Piaget's results to ensure that they are reliable; that the same age-wise variations in behaviour can be found using the same procedures. Whilst the Piagetian stages putatively form useful benchmarks of achievement, there has been little attempt to assess whether these are the result of behavioural change or cognitive change.

Inhelder & Piaget did not take into account whether performative factors (e.g. the effect of repeated trials, feedback, error information) may have caused, or contributed to, variations in behaviour. In the Piagetian theory, performative factors can only produce behavioural changes that are narrow and non-general. Cognitive change cannot result from training but only through the construction of cognitive structures from existing ones. However, without checking whether performative variables have an influence on variations in seriation performance, the inference that these variations are indexical of changes in cognitive competence is only a weak inference. Given subjects in a state of cognitive readiness for change (as per Vygotsky's zone of proximal development), timely intervention might result in subjects being able to take advantage of this readiness to produce behavioural changes of a lasting nature.

In order to have strong evidence that seriation is indexical of cognitive change it is necessary to provide intervention to individual subjects to see whether or not assistance can result in behavioural change. Further, in order to ensure that change is robust and reliable, subjects need to be retested. If seriation cannot be confirmed as an index of changes in cognitive competence then there is little point in further extensions of the clinical method.

If we find that our results agree with those of Inhelder & Piaget, then we may proceed with the question of whether or not intervention will change the stage-like pattern of variation in behaviour. Young children's failure

to produce a correct copy of the seriation model may not be due to an inherent inability to seriate. Failure to seriate may be due to a shallower problem of misunderstanding the task or problems with being aware of errors, locating errors and repairing them (see Sonoda, unpublished manuscript). Such problems might be easily ameliorated by simple intervention: giving subjects opportunity to review their productions, prompts, hints, etc. Subjects might then use information gained from this experience to succeed at seriation unaided (as per Vygotsky's notions of internalization and the zone of proximal development).

We use an error-motivated intervention procedure. Rather than terminating the task when there is a failure to correctly produce a copy of the model, children are given an opportunity to review their productions, locate their errors and repair them. To see whether subjects are aware of their own errors, they are encouraged to evaluate their productions in a direct perceptual manner by comparing the state of their own construction with the model (state-based feedback). After subjects have been made aware of their default, they are assisted in the location and repair of errors through a series of progressively stronger clues. If they manage to achieve a correct solution with intervention, they are retested to assess the uptake of assistance. Progressively stronger levels of assistance makes it possible to assess how much assistance subjects require in order to succeed --- how "near the mark" they are with respect to a self-guided solution.

Finally, this work centers on the identification of reasons for failure. We ask whether the same error factors are at work throughout the developmental range, or do children fail for different reasons at different stages of development. Error awareness, error location and repair are kept separate so that it is possible to diagnose the source of the default with respect to these three processes. By providing intervention it is possible to gain a further window onto change in seriation performance which is more fine-grained and long-term in time scale. If intervention does not result in any changes, or changes which are merely temporary, then it may be (putatively) concluded, as per Inhelder & Piaget, that performative factors have no role in the development of seriation. However if intervention results in changes in performance, which are not merely temporary, it may be concluded that performative factors do have a role in the development of seriation. If change is rapid, then it may be concluded

that subjects were in a state of readiness and required an opportunity for implementation; that readiness plus opportunity are the necessary and sufficient conditions for change. If change is slow, the subject may not be said to be in a state of readiness. In this case we are given a further window onto change. We can use the analysis of error information gained through intervention as an index of resistance points to change.

EXPERIMENT 1

Subjects

11 three year old and 27 four year old children at the Edinburgh University Psychology Department Nursery, 17 five year old children at Bruntsfield Primary School, and 23 six year old children at Sciennes Primary School.

Apparatus & Materials

2 sets of 10 wooden blocks. All blocks were 39 mm high and 39 mm deep with lengths 19 mm, 32 mm, 45 mm, 58 mm, 71 mm, 84 mm, 97 mm, 110 mm, 123 mm, 136 mm. When blocks were arranged as a model for subjects to copy, they were separated by approximately the width of a block (39 mm). Subjects were individually tested in a small quiet room seated at a table with the experimenter seated opposite. The table in the nursery testing room was 3.5 feet in diameter and the school testing room table was 2.5 by 3 feet.

Design

The experimental design is shown schematically in Figure 2.1. Each experimental session was divided into three parts: seriation, intervention, and retest. First subjects received a standard seriation test as per Inhelder & Piaget. If subjects spontaneously succeeded, they immediately received a second seriation test. If subjects spontaneously failed, they received intervention. If success was achieved through intervention, subjects received a second seriation test. If success was not achieved through intervention, the session was terminated after 20 minutes had elapsed.

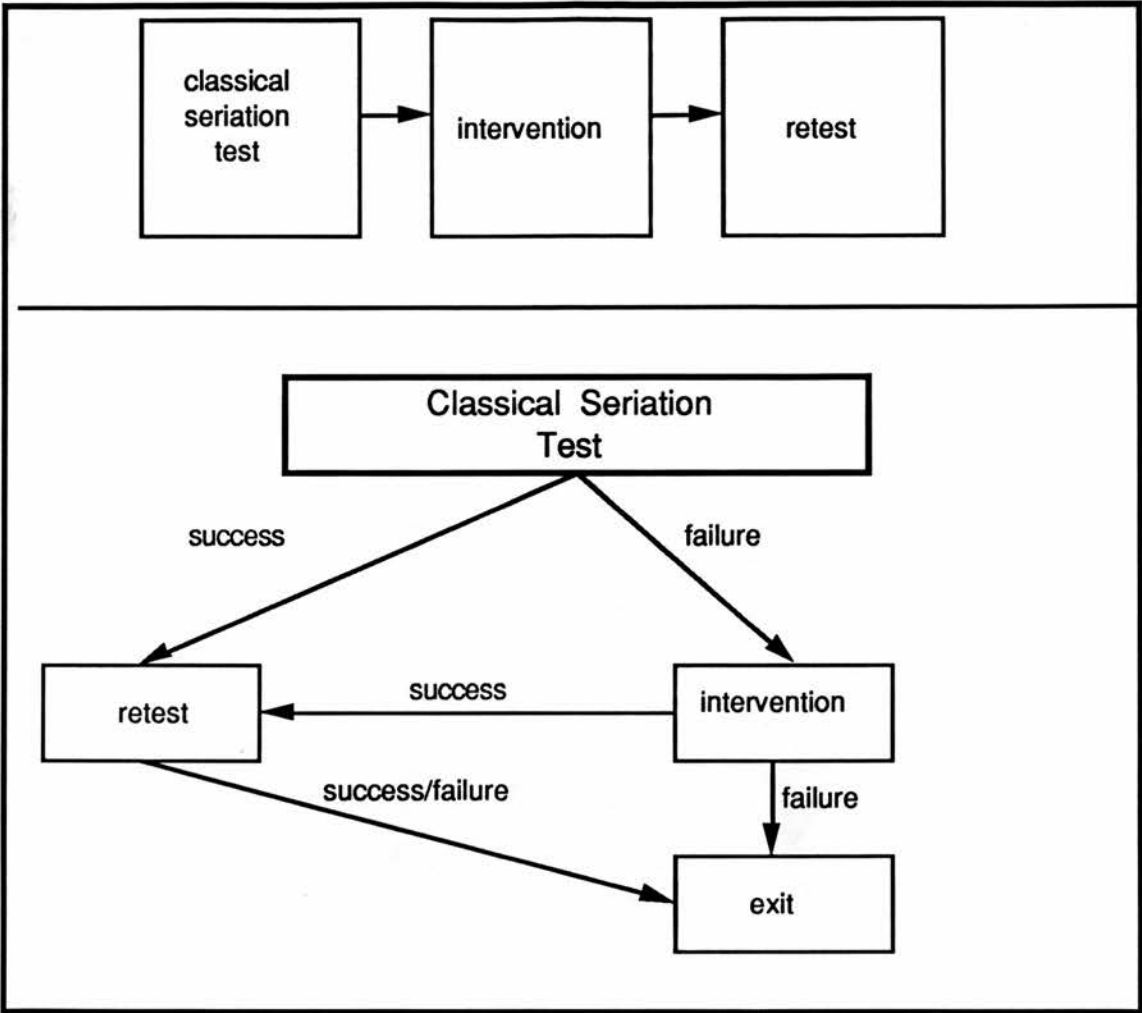


Figure 2.1: Design

Procedures

Basic Seriation Task. The experimenter asked subjects to watch very carefully while the set was seriated “because after I’m finished making mine, we’ll see if you can make your blocks just like mine.” Each block had a white dot to designate its bottom and there was a taped line on the table to indicate where blocks should be placed. The experimenter drew subjects’ attention to this. Subjects’ blocks (10) were arranged in an unordered pool (no block is on top of any other block) away from the area where they place them for seriation. The experimenter’s completely seriated set were placed in front of subjects so that they could look at (but not touch) them at any time. Once the model was completed, subjects were asked to “make one just like mine” with their blocks.



Figure 2.2: A subject performing the seriation task

Intervention Procedure. After subjects had finished their arrangements, the experimenter told them to align their blocks along the tape (if they had not already done so) and ensured that all the blocks were placed with their white stickers at the bottom. The experimenter then asked some questions. First the experimenter asked subjects if their arrangements were “just like” the model. (a) If subjects answered that they were the same, when they in fact were not the same, the experimenter disconfirmed this and encouraged subjects to change their minds. The experimenter then offered a weak cue as to the location of the error. For example, “Look at my blocks. They are smooth along the top and yours are bumpy. Can you show me the bumps?” If subjects failed to locate an error, then the experimenter offered a stronger error location cue. For example, “I see some bumps here in the middle.” If subjects persisted, the experimenter pointed out a particular error. For example, “This block is out of place.” Once an error had been located, subjects were given an opportunity to spontaneously repair. If subjects failed to make a repair, the experimenter requested that subjects make a repair. For example, “Can

you fix the bumps?" If subjects persisted, the experimenter offered a weak repair cue. For example, "Why don't you try moving a block." If subjects persisted, the experimenter offered a stronger repair cue. For example, "Why don't you try moving a block over this way." If subjects persisted, the experimenter made the repair. Subjects were once again invited to review their productions and the experimenter continued offering assistance in this manner until the array was completely corrected or until 20 minutes had elapsed. (b) If subjects failed to answer the question or said they don't know, then the experimenter offered assistance and asked subjects to try and repair the arrangement, as above. (c) If subjects answered that their arrays were different from the model (whether it was or not), the experimenter gave subjects an opportunity to spontaneously locate an error. At this point an opportunity was given for a spontaneous repair. If there was no spontaneous repair, the experimenter requested that subjects make a repair and continued offering weak repair cues (and if required, stronger repair cues) until the repair was made. If subjects did not spontaneously locate an error then they were asked to do so. For example, "Where is yours not like mine?" If subjects still failed to locate an error, the experimenter gave a weak cue and continued as above. (d) If subjects answered that they were the same, and indeed they were, the experimenter went immediately to the retest.

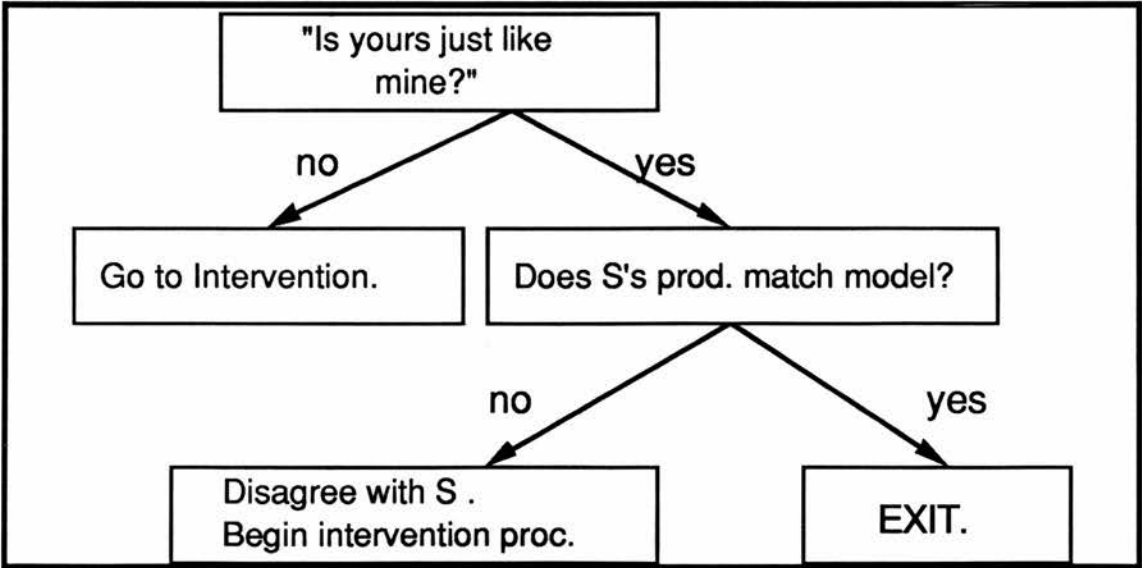
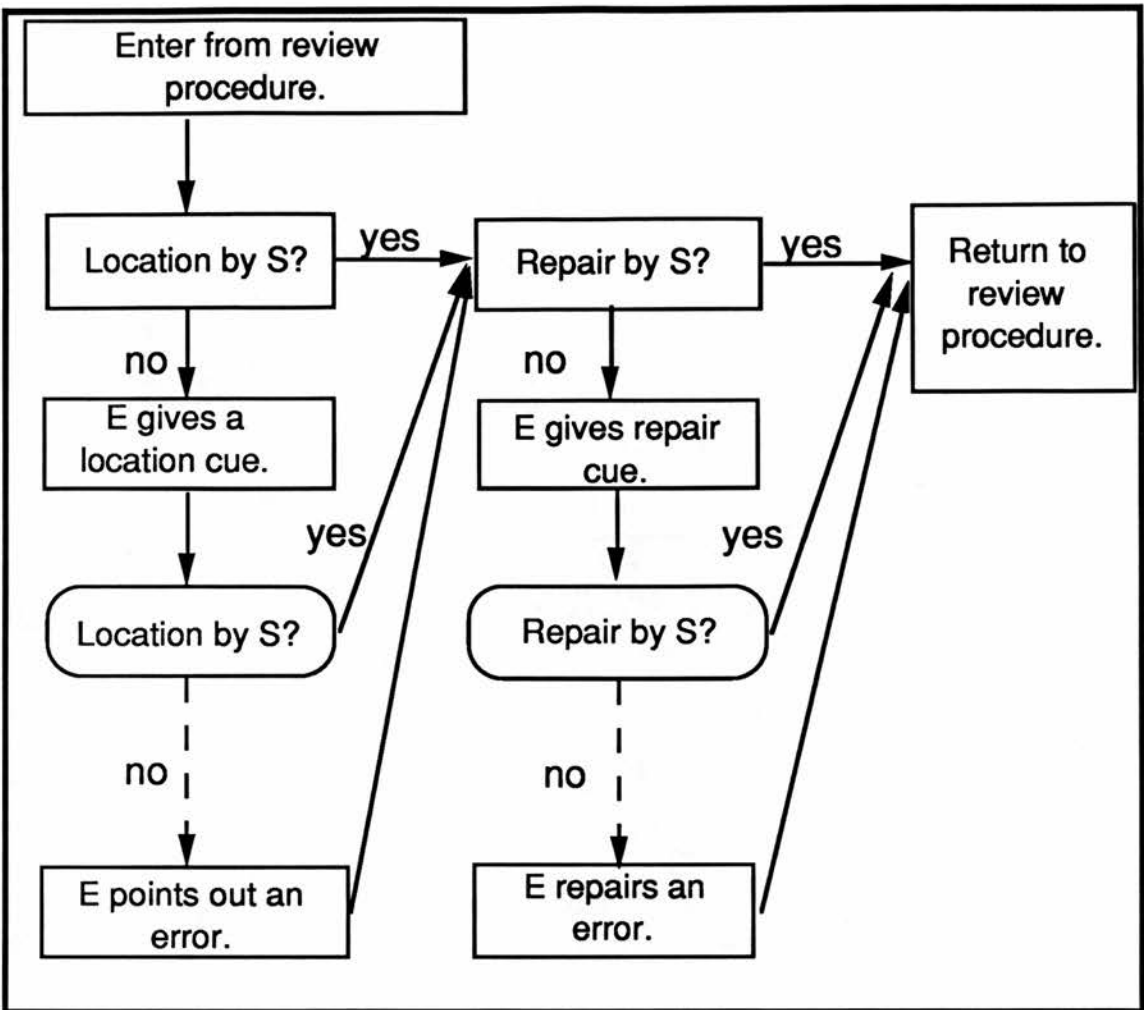


Figure 2.3: Review Procedure



Key to Levels of Intervention		
Level	Location	Repair
Level 1	Requested Location	Requested Repair
Level 2	Weakly Cued Location	Weakly Cued Repair
Level 3	Strongly Cued Location	Strongly Cued Repair
Level 4	Location by Experimenter	Repair by Experimenter

Figure 2.4: Intervention Procedure

Seriation Retesting. A second seriation test was given immediately after subjects had completed the intervention procedure above. If subjects had not achieved a correctly seriated set of blocks during the intervention procedure, they did not receive a second seriation test and exited the study at this point.

The experimenter congratulated subjects on performance in the first task and said “Let’s play another game where you make your blocks just like

mine.” The experimenter returned the subjects’ blocks to the pool and asked subjects to “make it just like mine again.” When subjects indicated that they had finished with their productions, the experimenter asked “Is yours just like mine?” The subject’s response was noted and the session was completed.

The tasks described above were administered in one session of approximately 20 minutes. All sessions were video taped and portions were audio taped as well.

Results

x

AGE		3	4	5	6
no. of subjects	I & P	0	15	34	32
	DN	11	27	17	23
FAILURE AT SERIATION	I & P	##	100%	79%	41%
	DN	100%	56%	71%	30%
"TRIAL & ERROR" SUCCESS	I & P	##	0%	12%	25%
	DN	0%	37%	6%	39%
"OPERATIONAL" SUCCESS	I & P	##	0%	9%	34%
	DN	0%	7%	24%	30%

81
67

57% p 1114 - weaker category =

Table 2.1: Comparison with Inhelder & Piaget

Comparison with Inhelder & Piaget

Table 2.1 compares the present results with those of Inhelder & Piaget. Their similarity confirms the reliability of Inhelder & Piaget's results for a “first pass” at seriation. Although there is some difference in the performance of four year olds, this may be due to a sampling error either on the part of this study or Inhelder & Piaget’s. Note that Piaget tested only fifteen four year olds. Further, here “failure” (Inhelder & Piaget's stage 1A) has been collapsed with “uncoordinated sets” (stage 1B). This category was eliminated because its validity as a true “stage” in

competence in the development of seriation was doubted.¹ Given the modified table however, the similarity of the present results to those of Inhelder & Piaget confirms their reliability. In this way we can now be sure that we are working within a task domain which is pointing to real changes occurring over a long period of time. Having confirmed the long-term time scale for change we are now licensed to revise and extend the clinical method in order to see if seriation can be made more transparent. Because the Piagetian results still stand, there are points of reference within which to situate further, more fine-grained assays of change.

Breakdown of Performance

In order to get away from the strict classification of subjects by age, subjects were first divided into functional taxonomic groups. These groups were based upon subjects' performance. How these groups associate with age is explored later. Subjects were initially divided into two groups, *spontaneous success* and *spontaneous failure*, based on their performance on the first seriation task. Out of the total group of 78 subjects 33 (42%) succeeded spontaneously on both the initial seriation test and the subsequent retest and 45 (58%) spontaneously failed. Spontaneous success was further broken down into three functional taxonomic groups: principled, systematic, and asystematic. These were based upon the style of successful performance. *Principled success* involved the perfect execution of the "operational method" in one smooth serial production; that is, the

¹We introduced a modification to these results however, eliminating "small uncoordinated sets" as a stage. Because partial collections are likely to be produced by purely random means, given the inbuilt constraints of the task, they have dubious significance as a stage in cognitive change. Piaget's notion of uncoordinated sets has to do with partial collections embedded within the parent set of ten items. Although Piaget saw this as a stage on the way toward full seriation, it is not clear whether partial collections have any significance for the subject or whether they are accidents of arrangement. The ten-item pool Piaget used in the seriation task is constrained such that it contains *all and only* the items required to reproduce the model. In a purely random production where such a pool is used and the requirement is that each item in the pool must be used, there is a high likelihood that the subject will get at least some of the items together in a monotonic set without any insight whatsoever. If uncoordinated sets can be produced by purely random means, then such behaviour does not reflect a benchmark of competence in seriation. There is no reason therefore, to include it as a stage in the development of seriation behaviour separate from failure.

subject selects the biggest (or smallest) block first, and then the next biggest, etc., until the set is exhausted. *Systematic success* involved a close approximation of this method with some trial and error. *Asystematic success* involved success using no obvious principle for selecting and placing blocks and requiring a great deal of trial and error. Of those subjects succeeding spontaneously, 13 (17%) succeeded by principled means, 18 (23%) succeeded by systematic means, and 2 (3%) succeeded by asystematic means.

Spontaneous failure was also divided into three functional taxonomic groups: transitional success, short-lived success, and robust failure. These were based upon the subjects' response to intervention. *Transitional success* involved success achieved with assistance provided in the intervention phase and subsequent unassisted success in the seriation retest. *Short-lived success* involved success achieved with assistance provided in the intervention phase, and subsequent failure after assistance is withdrawn in the seriation retest. *Robust failure* involved failure despite assistance provided in the intervention phase. Of the 45 (58% of the total group) who spontaneously failed on the seriation test 14 (18%) were able to succeed through intervention. However, a further 6 (8%) transitional successes resumed failure once assistance was withdrawn (short-lived success). 25 (32%) subjects of the total group failed persistently despite assistance provided.

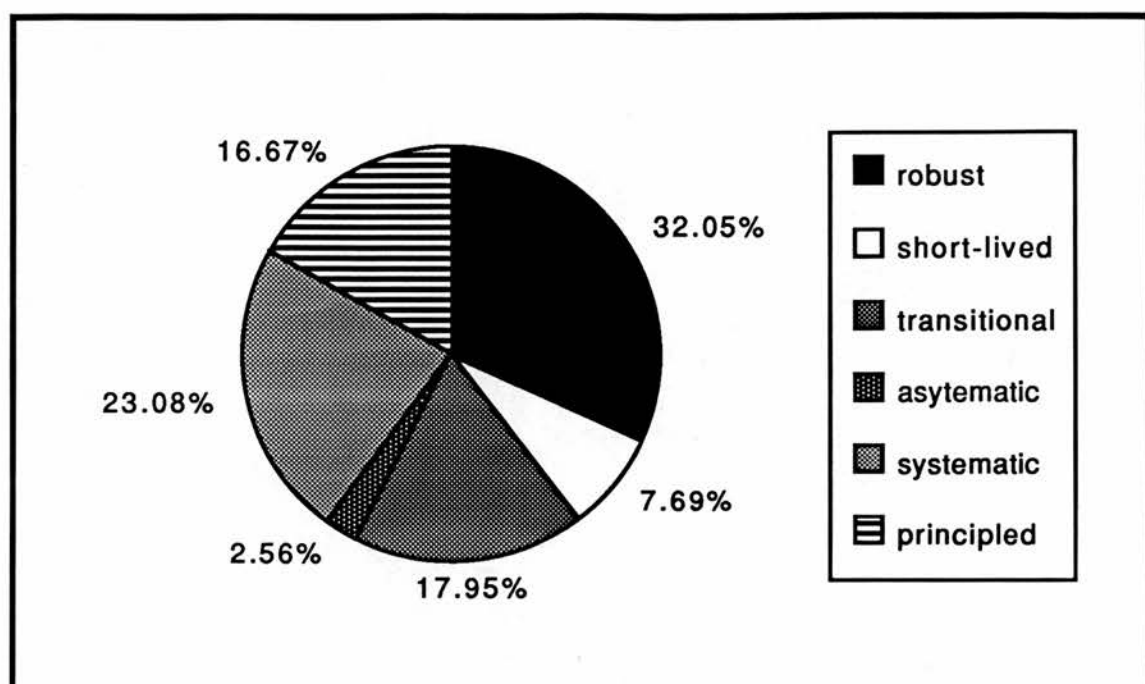


Figure 2.5: Breakdown of performance by functional taxonomic group.

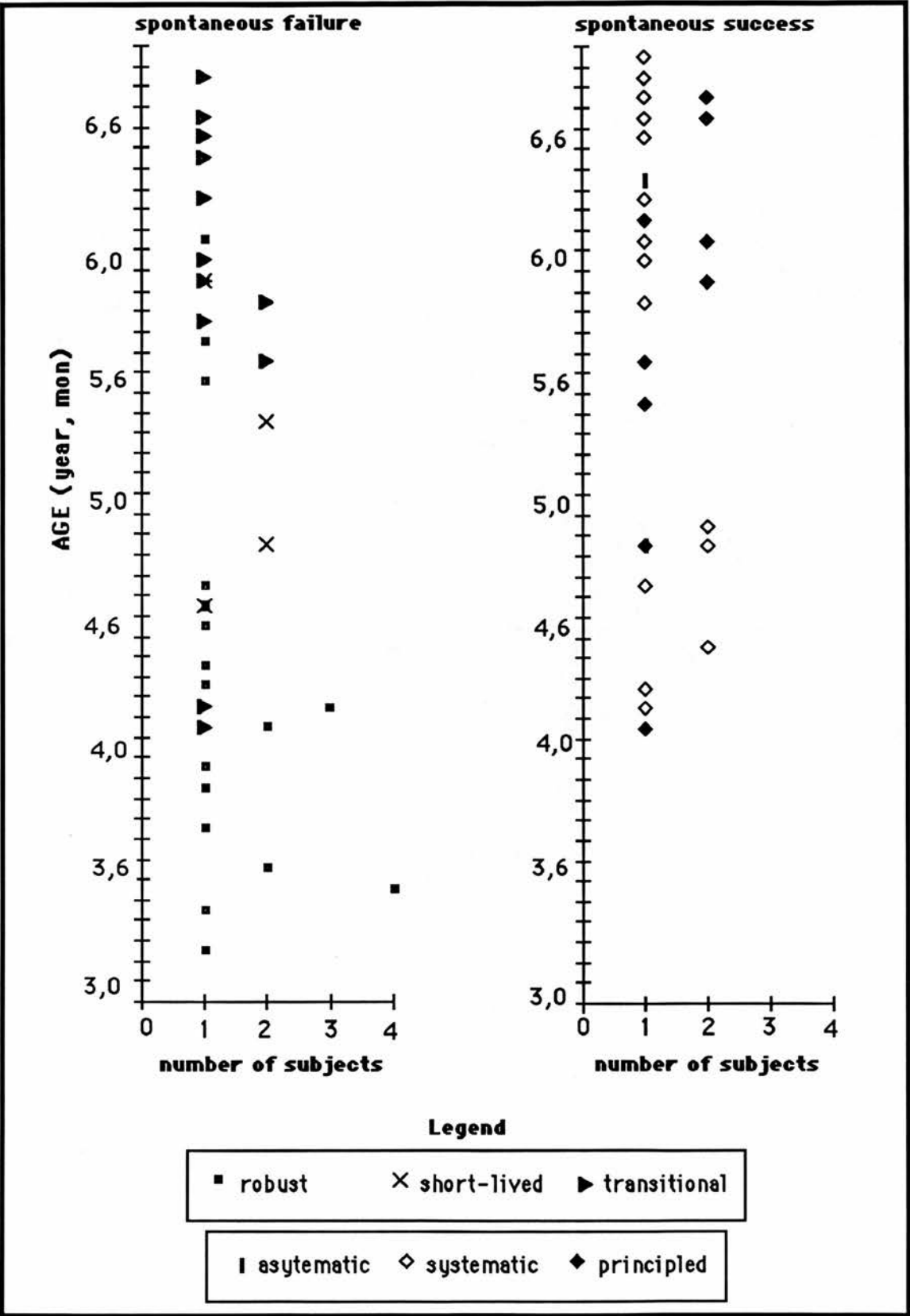


Figure 2.6: Age distribution of functional taxonomic groups

Age Distribution

The distribution of functional taxonomic groups by age is shown in Figure 2.6. Whilst spontaneous failure was found across the entire age range, success by any means, including with assistance, was not found below age four years. This age range belonged exclusively to the robust failure group. After age five years however, robust failure was uncommon (there were only three further cases of robust failure between five-and-a-half years and six years one month). There were only a few cases of short-lived success, however these were concentrated between ages four-and-a-half and five-and-a-half years. All but two transitional successes were found upwards of age five-and-a-half years (mean age five years, ten months), a time when their peers may be succeeding spontaneously. All subjects over age six years, one month who initially failed could become successful through intervention.

Spontaneous success ranged from ages four years, one month to age six years, ten months (the oldest child in the study). There were only two asystematic success subjects, one at four years, ten months and one at six years, four months. Systematic and principled success had similar age ranges. Systematic success subjects were found between ages four years, two months and six years, ten months. Principled success subjects were found between ages four years, one month and six years, eight months. The mean age for systematic success subjects was five years, six months. The mean age for principled success subjects was five months older at five years, eleven months.

Comparison of subject status pre- and post- intervention

Table 2.2 shows subjects' status before and after intervention, according to the modified Inhelder & Piaget classification scheme (as per Table 2.1 above).

AGE		3	4	5	6
no. of subjects	<i>DN</i>	11	27	17	23
	<i>post</i>	11	27	17	23
FAILURE AT SERIATION	<i>DN</i>	100%	56%	71%	30%
	<i>post</i>	100%	48%	35%	4%
"TRIAL & ERROR" SUCCESS	<i>DN</i>	0%	37%	6%	39%
	<i>post</i>	0%	44%	35%	43%
"OPERATIONAL" SUCCESS	<i>DN</i>	0%	7%	24%	30%
	<i>post</i>	0%	7%	29%	52%

Table 2.2: Results after intervention compared with previous results

For three year olds there was no change. For four year olds there was some reduction in failure as two subjects changed from failure in the first seriation test to "trial and error" success (asystematic) in the seriation retest. For five year olds the reduction in failure was dramatic with six subjects changing from failure to success. Of these six subjects, five succeeded by "trial and error" (systematic) on the retest and one succeed by the "operational" method. For six year olds the reduction in failure was also dramatic. Of seven subjects failing on the first test, only one did not achieve success with intervention. Of these six subjects, five succeeded by "trial and error" (systematic) on the retest and one succeed by the "operational" method. Further, there were five six year old subjects who succeeded by "trial and error" on the first test (four systematic, one asystematic), but had upgraded their performance to the "operational" method on the retest. There was one subject who did the opposite, changing her performance from "operational" in the first test to trial and error (systematic) in the second test.

Intervention Error Analysis

Trials to criterion or exit. All subjects who failed spontaneously participated in the intervention procedure. Subjects in the transitional success group required an average of 3 trials (intervention cycles) before completing a correctly seriated production. Similarly, subjects in the short-lived success group also required an average of 3 trials. Subjects in the robust failure group however, received an average of 13 trials before the session was terminated.

Error types and depth of intervention. Recall that in the case of an incorrect production, subjects were first required to answer a question about whether their productions matched the model or not, then they were given an opportunity to specifically locate and repair their errors. If the production still did not match the model, the procedure was repeated. Therefore, following this procedure, there was an opportunity for subjects to make three different types of error per trial: registration, location, and repair. *Registration errors* involved the subject's failure to register global disparity, between the production and the model as evidenced by an incorrect response to the question "Is yours just like mine?" in the review procedure. *Location errors* involved the subject's failure to locate particular disparities between the production and the model. *Repair error* involved the subject's failure to repair an incorrect production.

Because subjects were given the opportunity to spontaneously locate and repair errors, location and repair errors could be further examined in two contexts: *spontaneous* and *prompted*. In the case of prompting, the "depth" of the prompt required was noted. Recall that there were four levels of prompting in the intervention procedure. Firstly, the subject was prompted to locate or repair. Secondly, the subject was given a weak location or repair cue. Thirdly, the subject was given a strong location or repair cue. Finally, a location or repair was made for the subject.

Table 2.3 shows, by group, the percentage of total trials in which subjects failed to register global disparity, as well as the percentage of total trials in which subjects failed to *spontaneously* locate or repair errors. Table 2.4 shows, by group, how depth of intervention was distributed across trials (the percentage of trials in which each depth of intervention prompting was required). The top panel shows locations depths and the bottom panel shows repair depths.

GROUP	REGISTR.	LOCATION	REPAIR
TRANSITIONAL	56%	78%	68%
SHORT-LIVED	62%	76%	86%
ROBUST FAIL.	89%	95%	55%

Table 2.3: Error Type

LOCATION

GROUP	SPONT.	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
TRANSITIONAL	22%	12%	56%	7%	2%
SHORT-LIVED	24%	5%	52%	10%	10%
ROBUST FAIL.	5%	8%	39%	17%	32%

REPAIR

GROUP	SPONT.	LEVEL 1	LEVEL 3	LEVEL 2	LEVEL 4
TRANSITIONAL	32%	59%	7%	2%	0%
SHORT-LIVED	14%	24%	38%	10%	14%
ROBUST FAIL.	45%	22%	17%	4%	12%

Table 2.4: Depth of Intervention

Transitional success subjects

Although transitional success subjects had considerable difficulty registering their mistake, the greatest source of error was failure to spontaneously locate specific disparities between their own productions and the model. This error was committed in 78% of trials. The registration of global error and location of specific errors are of course related. Subjects are likely to have even greater difficulty specifically locating errors if they do not register the mistake in the first place. Indeed all three groups showed a higher failure rate for spontaneous location of specific error than for global registration of error. Typically transitional subjects required "weak" (Level 2) prompting for error location. This depth of error location prompting was required in 56% of trials. Only a small percentage of "request" (Level 1) prompting was required (12%), as well as only a small percentage of the "deepest" forms of prompting (7% and 2% for Levels 3 and 4 respectively). The problem of error location surmounted, error repair came more easily. Transitional subjects typically required only the shallowest form of prompting, "request" (Level 1) prompting. This depth of prompting was required in 59% of trials. However, there was a considerable amount of spontaneous error repair (32%) as well. It is likely that there would have been more spontaneous

error repair, however considering the "break" in manipulating blocks called for by questions about error location, subjects may have been waiting for the experimenter's "permission" before moving blocks around again. Indeed, spontaneous and "request" repairs together accounted for 91% of repairs.

Short-lived success subjects

The short-lived success group had error registration and location difficulties similar to that of the transitional success group, registration errors occurring in 62% of trials and locations errors occurring in 76% of trials. However, the most frequent source of error for short-lived success subjects was failure to repair errors. This error was committed in 86% of trials. In this way, difficulty seemed to "mount up" for subjects. Just as specific location was likely to be more difficult if there was already trouble in globally registering mistake, for these subjects repairing errors seemed to be even more difficult. As found with transitional subjects, "weak" (Level 2) prompting for error location was most typically required, accounting for 52% of error location prompts. Again there was little call for prompting at Levels 3 or 4. Compared with transitional success subjects, short-lived success subjects typically required deeper repair prompts. "Weak" (Level 3) prompts were required most (38% of the time), although spontaneous repair occurred 13% of the time and "request" repairs occurred 24% of the time.

Robustly failing subjects

Robustly failing subjects had a profound problem with registering globally disparity. These errors occurred in 87% of trials. The problem with specific location of error was even more acute, occurring in 95% of trials. Subjects typically required deep forms of intervention for error location. "Weak" prompting was required in most cases (39%), but experimenter locations (Level 4) were required about as often (in 32% of trials). "Strong" prompts were required in a further 17% of trials. Spontaneous and "request" locations only occurred in 5% and 8% of trials respectively. Different to the other groups, repair errors were less prevalent, occurring in only 55% of trials. Robustly failing subjects seemed to be keen to spontaneously repair their productions (45% of trials). Spontaneous and "request" repairs together accounted for 67% of repairs. However, subjects

in this group were less likely to make effective repairs than subjects in other groups. Productions were frequently made worse rather than better.

Results Summary

For the classical seriation task, results similar to those of Inhelder & Piaget were found, confirming their reliability. Having confirmed the indexical value of the seriation task as a pointer to real change, we were licensed to revise and extend the task in an attempt to make it more transparent.

Assistance offered through intervention was found to help some subjects succeed at seriation, whilst others did not benefit. Thirty-one percent of subjects who failed at seriation were able to take advantage of assistance and were able to succeed spontaneously thereafter. A further thirteen percent were able to succeed with assistance, but could not succeed spontaneously when assistance was withdrawn. Fifty-six percent of unsuccessful subjects were repeatedly unhelped by intervention and persistently failed. These results were found to be age-related. There was a clustering of transitional subjects between ages five-and-a-half and six. All subjects under age four failed persistently.

Error analysis revealed failure to specifically locate errors as a resistance point to success for all groups, but especially for the robust failure group. Transitional and short-lived success subjects required shallow forms of intervention before achieving success. Robust failure subjects required deep forms of intervention, yet were still unable to achieve success.

Discussion

We have confirmed that Inhelder & Piaget's results are reliable in that we found the same age-wise staged pattern of variance in seriation performance. This was a crucial first step in the investigation. Recalling that a general goal of the enterprise is to characterise transitions in competence, it was necessary to ensure that the chosen task domain was indeed indexical of transitions in competence. The putatively indexical nature of the Piagetian results --- a long development of age-related variations in competence, characterised as stages, made seriation attractive as a candidate domain for modeling transitions in competence. However there would be little point in pursuing this or any extended analysis if the stage-like pattern of change did not really exist. If Inhelder & Piaget's

results were idiosyncratic, then seriation would no longer be a suitable task domain. Further, it was important to ensure the reliability of Inhelder & Piaget's results, not only to license extended analysis, but also to ensure a baseline context of long-term cognitive development in which to situate it. Without this, extended analysis could be considered to be idiosyncratic to new procedures rather than being anchored in the classical Piagetian staged analysis.

Whereas Piaget looked at how different age groups of subjects fell into stages, we moved away from categorising subjects by age. Instead subjects were grouped by functional category to see how these declared themselves in terms of age. An asymmetry between success and failure was found, indicating that developmental factors are at work in changes in seriation performance. That no success was found below age four indicates that this may be a significant lower boundary for seriation. Similarly, the paucity of unsuccessful subjects over age six may indicate the upper bound of failure.

Whilst Inhelder & Piaget's results still stand, some subjects were able to improve their performance through intervention. The permanence of this change in performance for some subjects indicates that intervention can result in real changes in subject competence. However, this was not the case for all subjects. For some subjects improvement was only temporary. This would indicate that for these subjects there was no real change in competence, rather a temporary behavioural adjustment. That some subjects did not improve their performance, or changed only temporarily, indicates that whilst performative factors may be able to account for some variance in seriation performance, it is not the only factor. The six months between age five-and-a-half and six were where the highest concentration of transitional subjects was found. This is a likely candidate age range for a "zone of proximal development," i.e. an age range where performative variables are a factor in change.

More than giving subjects an opportunity to improve their performances, intervention gave us an opportunity to diagnose resistance points to success --- to open a window on failure at seriation which had previously been unanalysed. For transitional subjects, the problem was not knowing that they were wrong, but finding specifically where they were wrong. The small percentage of "request" (Level 1) prompting indicates that specific

error location was not just a matter of having to ask subjects to offer information which they had at hand. Similarly the small percentages of the "deepest" forms of prompting indicates that error location was not a profound problem. The high rate of "weak" prompting indicates a copying problem --- subjects needed to be invited to compare their productions to the model. Further, it was shown that this was a problem easily overcome. Once they were able to specifically locate errors, they quickly and easily repaired their productions. As their productions improved, they required less and shallower intervention. This indicates that for these subjects resistance points are not insurmountable, but rather just below the surface --- a further indication that these subjects are in a transitional phase between failure and success.

Short-lived success subjects had similar error registration and specific location problems to the transitional success group. However they had a further problem with repairing errors. It is this greater difficulty in repairing errors that seems to set short-lived success subjects apart from transitional success subjects. That "deep" forms of prompting were frequently required for errors to be repaired indicates that subjects may never have taken aboard the error repair process themselves. Although short-lived success subjects could achieve success with assistance, it is likely that when assistance was withdrawn, the combined problems of specifically locating errors and repairing them proved too much for subjects to achieve when performing "solo."

Robust failure subjects, on the other hand, were entirely resistant to success. These subjects persistently failed to recognise their error as such. Even when they were able to recognise error they were incapable of pointing out specific errors. Though most subjects were able to affect some sort of repair, it was suspected that these repairs were not error-informed attempts to approximate the model, but merely "acting" on the blocks as part of the "game." These findings indicate that such subjects had little in the way of error appreciation. Such profound limitations with respect to error awareness further imply that robustly failing subjects may have had a different understanding of the task than either adults or their older peers. That this sort of behaviour was typical of all subjects under age four may indicate four years of age as a developmental boundary between failure and the beginnings of success.

Young subjects' lack of error appreciation seemed to indicate that their understanding of the task was inadequate, that is, different from the task understanding of older children and adults. The intervention procedure was an attempt to better communicate the task by interacting with the subject --- calling attention to error, giving hints, and providing opportunities for error location and repair. Although these methods helped some subjects to improve their performance, as well as providing useful information to the experimenter, subjects under age four years did not show any improvement. The question is, can more effective methods be used to communicate task and error information to the subject, and also provide more information about seriation skills to the experimenter? This issue is discussed in the next chapter.

Chapter 3: New paradigms and procedures for the study of seriation

Communicating task and error information

Error information in the seriation task is usually communicated by state information, that is, subjects are required to monitor the state of their own productions by comparing them with a model in a direct perceptual manner. Error is signalled by (the detection of) a mismatch between the production and the model. A mismatch should serve as negative feedback, provoking the subject to make a repair. In the intervention procedure, subjects were actively encouraged to use state information as a means of detecting error. They were invited to look for mismatches and were provided with global and specific hints about their location. For young subjects however, this proved to be a difficult task even when assistance was provided.

However, it may be possible for children to use state information to detect errors, if other confounding factors were eliminated. Firstly, subjects' own productions may have a "privileged status" with respect to error; children are likely to be unwilling to think of their own productions as "wrong." In fact, blame for mismatches was likely to be put on the experimenter.¹ Secondly, in a seriation task subjects must construct their productions serially whilst simultaneously monitoring the state of the productions for discrepancy with respect to the model. It is possible that children can evaluate productions using state information, however the demands of serial production may have negative consequences for children's state evaluation abilities.

In a preliminary study (see Appendix A) we focussed on the question of whether children younger than age four years can recognize error in terms of state information alone when the demands of serial production have

¹One subject aged four years, seven months made a production using only eight blocks rather than ten. When asked, "Is yours just like mine?" he replied, "Yes, but I don't need these blocks. You gave me extras." He then handed them back to the experimenter.

been removed from the task. Pre-formed models and photographs of four different five-item series were used in order to clarify the issue of whether inability to register discrepancies between serial configurations was in part due to the "privileged" status of the subject's own production. The series used were monotonic ascending (5, 4, 3, 2, 1, where 1 is the largest block and 5 is the smallest), monotonic descending (1, 2, 3, 4, 5), a minimally degraded monotonic series (1, 3, 2, 4, 5), and a maximally degraded monotonic series (4, 3, 5, 1, 2). It was found that young subjects were not good at recognizing disparity between block constructions. Subjects did not register disparity between a monotonic ascending and a monotonic descending series. Only half of the subjects claimed that the maximally degraded series was different from either of the monotonic series. Less than half the subjects (38%) registered disparity between the minimally degraded series and either the monotonic series or the maximally degraded series.

These findings may be interpreted as providing further evidence that young subjects have difficulty using state information to detect errors. As attempts to assist children in the use of state information did not result in change in young subjects, it may be the case that state information is not the best way to provide feedback in seriation tasks. Further, the intervention procedure provided for feedback only once a production was completed. Although successful subjects should be spontaneously monitoring and evaluating their productions against the model as they are constructing them, the experimenter did not intervene to provide feedback or hints (that the subject should be paying attention to state information) whilst subjects were actually constructing their productions. By this time it may be too late for subjects to grasp what has gone wrong. A more direct and unambiguous way of communicating error information is needed. Since immediate feedback is more usual in learning situations, it may be possible that subjects could learn to seriate if given immediate feedback. Owing to difficulties with state-based error analysis, copying and comparing with respect to a model could be eliminated. Introducing direct and immediate feedback (in the form of a "bleep") would instantly communicate error to the subject in an unambiguous way.

A further issue with respect to task understanding is stimulus set size. How would subjects who otherwise fail perform if they were allowed to first build up from smaller sets to larger sets? Piaget's choice of ten items as the stimulus set size was designed to force operational solutions rather than perceptual ones. However, on what basis was ten items chosen? Is seven or eight items not enough to bring out an operational solution? It may be that ten items had a psychometric value for Piaget in that it was only at about age seven that children could control such a large number of items using the operational method. However set size was never really examined as a variable in seriation. Ten blocks is a large set size, perhaps too large for young subjects to control immediately. If subjects were gradually incremented from smaller sets to larger sets based upon meeting a success criterion, it would be possible to determine what number of items subjects can control at particular times and at which number of items they meet resistance. By parsing the problem for the subject in this way, it is possible for the experimenter to assess stimulus set size as a variable without the problems of smaller sets being embedded in a larger "goal" set.

Decomposing seriation skills

Although the focus thus far has been on subjects' errors and their diagnosis, there are further questions to be answered about the root causes of failure. Recall the decomposition of seriation into several components including serial order, ordinal computation, modeling (copying) and search. By this characterisation, success is due to the synthesis of components, and failure is due to the absence or non-integration of one or more components. In the classical seriation task component skills are compounded. Therefore, in the case of the failing subject, it is difficult to distinguish which skills are defaulting by looking exclusively at performance on the classical task. Consider especially the difficulty in discovering the loci of failure in robustly failing subjects (as above). Although analysis of their errors yielded some understanding of their failure (e.g. inadequate task characterisation) further analyses of behaviour are required to discover the causal factors which determine seriation default. In order to achieve this, new paradigms and procedures are required in order to partition the failure space. We require methods which allow us to locate which competences subjects who otherwise fail at

seriation (that is, as classically presented) have and which they do not have. Also, we need to discover which competences are "teachable" and which are not. In this way it is possible to discover which obstacles in the path to seriation success can be overcome by subjects and which are met with resistance.

In this thesis, focus is concentrated on the serial order component. Although any linear arrangement of items is legitimately a series, the study of seriation has focussed on a special case of serial control, that is, the serial control of items which are monotonically ordered by size. But what about children's serial ordering abilities *per se*? How would children perform if asked to construct a series of unrelated items? Because serial control of arbitrary series has been shown in nonhumans (e.g. Straub & Terrace, 1981; D'Amato & Colombo, 1988), it is likely to be a basic skill present in young children (McGonigle, 1987). However, there are limits on the length of arbitrary strings which can be managed by serial control alone (nonhumans have been found to control five items at most). If a higher means of control could be recruited to constrain the amount of search involved in the task, greater numbers of items could be accommodated, e.g. dimensional control, that is, taking advantage of ordinal size information in the monotonic size series. However, we have yet to examine children's serial control abilities using both arbitrary series and non-arbitrary series, such as the monotonic size series. Here there is also the question of whether or not a non-arbitrary, monotonic size series holds any special advantage, in terms of set size capacity, for young subjects.

New paradigms and procedures

In order to investigate the issues raised above, we employ training paradigms for the study of seriation, inspired by the decomposition (see McGonigle and Chalmers, in press, for a full report of new paradigms for the study of seriation components in isolation and coalition). As the focus here is on serial control aspects of seriation, we make use of two paradigms in particular. The *serial order paradigm* (see Straub & Terrace, 1981) is used here to study serial control *per se*, in this case, the serial control of arbitrary (colour) strings. The "*combined*" paradigm is similar to the serial order paradigm, however it demands the combined use of both sequential and seriational components (see McGonigle, 1987). It

involves the serial control of stimuli which are ordinaly related along a size dimension and is used to explore the interaction between serial and the ordinal components of seriation. Young children can be trained using these paradigms in order to investigate the teachability of seriation skills and to explore resistance points to success. The paradigms are represented schematically in Figure 3.1 (note that different patterns stand for the different colours in the actual task).

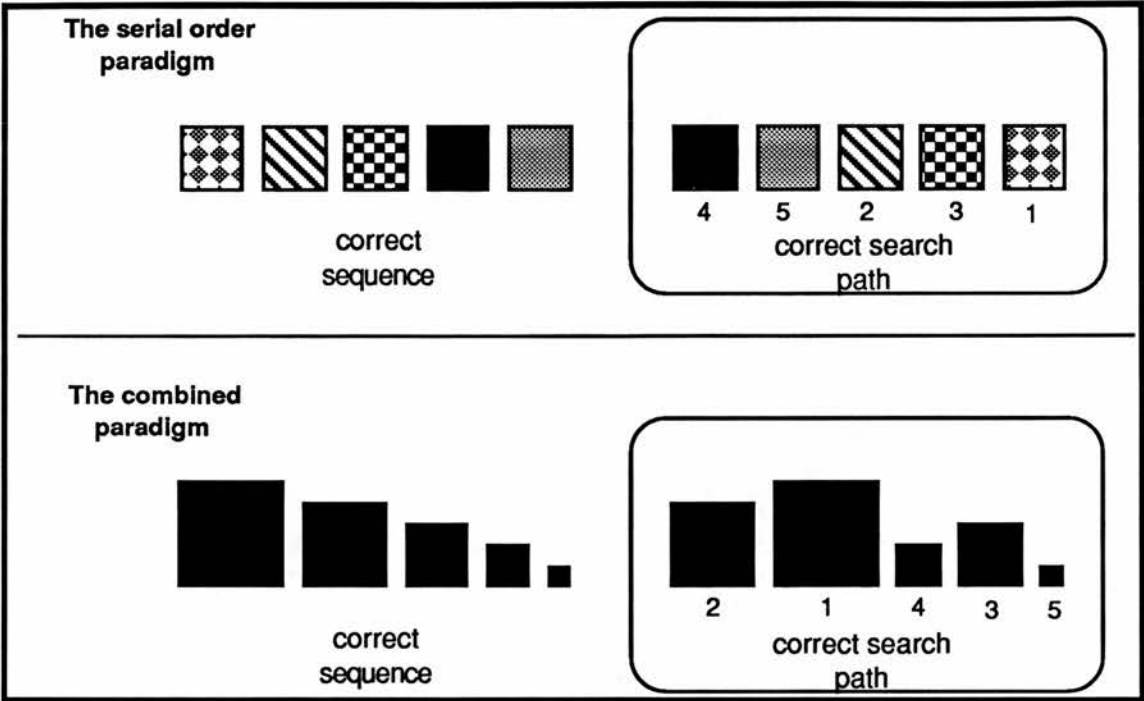


Figure 3.1: The paradigms

The paradigms are implemented using a particular apparatus, a computer equipped with a touch-sensitive screen. The subject responds by touching stimuli generated by a computer and displayed on the screen. Touching a stimulus results in the sounding of a registration tone. The subject is required to discover, by touching one stimulus at a time, which sequence of touches is correct. Incorrect choices are immediately signalled by the sounding of a negative feedback tone (after the registration tone), the trial is terminated and the subject must begin again. At the end of a correctly completed series of touches, a positive feedback tone is sounded and a new trial is presented. The position of objects changes from trial to trial to avoid clues from spatial position.



Figure 3.2: A subject performing the serial order task on the touch screen

In addition to immediate feedback, an important feature of training is parsed input. Subjects begin with a small number of stimuli and are required to repeat the correct sequence of responses in successive trials until they reach a performance criterion. At this point an additional item is introduced and the subject begins again. Further to immediate feedback, indirect feedback is provided by a line at the bottom of the screen which becomes longer as the subject approaches the performance criterion. In this way subjects can keep track of their performance throughout the session.

There are several advantages of using these paradigms as implemented on the touch screen apparatus. Firstly, it is possible to investigate whether or not seriation skills are trainable in young children who spontaneously fail in classical versions of the seriation task. Secondly, the use of immediate feedback allows for error information to be communicated directly. The problems of copying from a model and state-based error analysis are eliminated. The direct communication of error information should also

serve to improve subjects' understanding of the task. Thirdly, parsed input also assists task understanding by keeping the task simple at first and increasing difficulty gradually in terms of set size. Further, parsed input allows for set size to be explored as a variable. In this way it is possible to discover what set size capacity subjects have and where their resistance points are. Finally, it is possible to explore the failure space with respect to serial control and dimensional control, i.e., does the monotonic series hold any special advantage for young subjects.

The touch screen itself is important to the implementation of these paradigms, because it allows for the rapid presentation of many trials in a session. This is important considering the need to maintain the interest of young subjects over the course of training. Also, the automatic provision of feedback and other task information means that the experimenter can keep verbal instruction to a minimum. This is important in avoiding confusion about the task in young subjects whose linguistic skills may be very basic. Finally, because the experimental situation is controlled by a computer, data can be collected automatically and accurate measures of reaction time can be taken.

Although relevant to many subject groups, the issues raised here are especially interesting with respect to subjects under age four whose failure seems otherwise complete and impenetrable. The identification of the root causes of failure in these subjects was hampered by subjects' persistent and complete failure at seriation. There seemed no way to further analyse the dimensions of their failure within the classical paradigm. Finding the baseline competences these subjects have would be of particular value in the analysis of the transition from early failure to later success. These subjects were targeted for in-depth intervention. The results of this investigation are reported in Chapters 5 and 6. Before entering into an in-depth investigation however, it is first necessary to examine the feasibility of using the touch screen with young children. At this time, we know of no investigations using such an apparatus with young children. However, studies of serial order using similar methods have been carried out with nonhumans. An investigation of serial order in young children using the touch screen is reported in the next chapter.

Chapter 4: Assessing serial ordering skills using the touch screen

Rationale

Whilst there is a large existing literature on the serial ordering abilities of pigeons (Straub & Terrace, 1981; Terrace, 1987; 1991) and monkeys (D'Amato & Colombo, 1988; 1989; 1990), to the best of our knowledge, there has never before been a similar investigation of serial ordering abilities in young children. The principle aim of this investigation therefore, is to examine serial learning *per se* in young children by using serial order training and test procedures which have previously been used with nonhuman subjects.

As this study represents the first use of the serial order paradigm with young children, there are many questions about how these procedures may best be implemented with child subjects. A particular feature of the animal studies is training using immediate feedback over the course of many trials. Because few child studies feature such long-term training, the question remains how children will respond to a training study which features a large number of trials. Also featured in this study is a new apparatus, a computer equipped with a touch-sensitive screen. No experimental studies using such an apparatus with children are known to us at this time. Although there are many advantages to using the touch screen (as discussed in the previous chapter), we have yet to ascertain how children will respond to such an apparatus. Therefore an important goal of this investigation is to explore the conditions of serial order training and testing, with a view toward optimizing these conditions for child subjects. In this way, the general features of training and testing may be preserved throughout the totality of the procedures which follow in later investigations (see Chapters 5 and 6).

In particular we examine test procedures used to evaluate serial learning ability. Both animal studies put a great deal of store in results obtained from test trials in which the subject is required to skip over "missing" items in the series (e.g. BD). Success at such "gap-skipping" trials,



especially those which do not include the initial and terminal items (A and E), has been interpreted as indicating that subjects have some kind of effective internal representation of the series. Pigeons show performance at chance level on interior test pairs (BC, BD, CD) and triplets (BCD) whilst monkeys show a high level of performance on these subseries. These results have been interpreted as indicating that monkeys have an effective internal representation of serial order, whereas pigeons use a simpler means of reporting the series based on the special status of the initial item (A) and the terminal item (E). However in both studies the subjects are trained in a "forwards" direction. Subjects were presented the stimuli as they were to be reported in the final sequence, i.e. the first item first, the second item second, etc. It is possible that this means of training may have biased subjects toward skipping missing items in the test phase without necessarily indicating that they were using an effective representation of the series ABCDE. Would subjects be successful at such "gap-skipping" test trials were they to be trained in a different way, for example, if subjects were presented the stimuli in the opposite direction to how they were to be reported in the final sequence, i.e. last item first, fourth item next, etc.? Straub et al. (1979) attempted to use such a "backwards" training condition with pigeons, but abandoned it after the pigeons' lack of success.

In this study we investigate both forward and backward methods of training for two reasons: firstly, to discover which method is the best way for the child to be trained, and secondly, to determine the effect of training direction on performance in "gap-skipping" test trials. In addition to "gap-skipping" test trials used in animal studies we propose and investigate the feasibility of using new test procedures which focus on working memory demands.

Finally, because procedures were preserved with those used in animal studies (as far as was possible), the present data can be compared with existing results for pigeons and monkeys. In particular we ask: what is the status of the child as a serial performer *vis a vis* the pigeon and the monkey? Terrace and D'Amato & Colombo have obtained different results which have been interpreted by D'Amato & Colombo as indicating species differences. In particular, four- to five-item series seem to be taxing the pigeon's memory span, whilst monkeys have much less difficulty

learning a five-item series (ABCDE). D'Amato & Colombo (1989) have argued that such quantitative differences in performance betray qualitative differences in serial ordering ability. We explore how the child fits into this scenario.

This experiment has three main goals. The first is to examine serial ordering abilities in young children. The second is to test the feasibility of using new apparatus and new paradigms which have never before been used with young children and to establish procedures which can be preserved for further studies of seriation abilities within the thesis' experimental programme. Finally, this investigation seeks to compare serial ordering abilities in young children with existing data for pigeons and monkeys.

EXPERIMENT 2

Subjects

12 children from the Edinburgh University Psychology Department Nursery, aged from 3;9 to 4;7 (years, months) served as subjects. This age group of subjects was chosen for pragmatic reasons. Firstly, since daily training sessions were required, it was necessary to have daily access to subjects. This was possible using the Psychology Department nursery facilities. Further, as an important aim of this study was to assess the feasibility of new apparatus and new procedures, we sought subjects who might reasonably be expected to succeed, but not without some measure of difficulty. Based on results from the extended clinical assay, seven subjects who were likely, but not certain, to fail at seriation were chosen, that is, subjects older than four years (the oldest child in the nursery at this time was 4;7). Five additional children under age four years (candidates for the robust failure category) were chosen in order to further assess procedures with subjects more likely to fail.

Apparatus & Stimuli

Subjects were individually tested in a small quiet room. Subjects were seated on a height-adjustable chair at a table in front of a computer screen with the experimenter seated beside the subject. The programming of trial events, stimulus presentation, and data recording were controlled by a BBC microcomputer equipped with a touch-sensitive screen. A touch was

registered by the interruption of infrared beams projected horizontally and vertically across the screen. Because the spatial resolution of the beams was low, it was necessary to pre-train subjects to touch the screen appropriately (see pre-training procedure below).

The stimuli consisted of five 3.5 cm by 3.5 cm squares of different colours (red, green, yellow, blue, pink). The stimuli were displayed in a horizontal line across the middle of the screen, separated by 1.5 cm. Five items was chosen as the set size for three reasons. Firstly, five items was the maximum set size used in the animal studies. Secondly, five is the minimum set size in which there is an interior nonadjacent pair (BD), a crucial test item in "gap-skipping" transfer trials. Further, in a preliminary seriation study (see Appendix B), it was found that two-thirds of subjects under age five years who could not seriate ten items, could not seriate as few as five items either. In this respect it was expected that five items in an arbitrary set would be a challenge for young subjects in this study.

Design

The serial order paradigm used to assess the serial ordering ability of children is derived from Straub & Terrace's (1981) temporal order paradigm designed for use with pigeons. The subject must select (using a touch screen) a particular sequence of unrelated (in any principled way) objects, e.g. a five colour sequence. To prevent any cues from spatial position, each item changed location from trial to trial. Only the sequence remains invariant. Subjects were assigned to one of six colour sequences. Subjects were further assigned to one of two conditions: forward training and backward training. In the forward training case, subjects were presented the stimuli as they were to be reported in the final sequence, i.e. the first item first, the second item second, etc. In the backward case subjects were presented the stimuli in the opposite direction to how they were to be reported in the final sequence, i.e. last item first, fourth item , etc. The experiment was divided into six basic parts: pre-training, training, verbal probe, over-training, test phase I, and test phase II.

SUBJECT	AGE	CONDITION	STRING
BARRY	4,7	forward	red green yellow blue pink
JANE	4,6	forward	pink blue yellow green red
PAULA	4,6	backward	red green yellow blue pink
CHRISTINA	4,6	backward	red green yellow blue pink
FINLAY	4,6	forward	pink blue yellow green red
BEN	4,3	forward	red green yellow blue pink
LOUISE	4,1	forward	green pink blue red yellow
AMY	3,11	backward	yellow blue green red pink
LEWIS	3,11	forward	yellow blue green red pink
ANNA	3,10	backward	blue yellow red pink green
KIRSTY	3,9	forward	pink red yellow green blue

Table 4.1: Subjects' condition and string assignment

Procedures

Pre-Training. The subjects were introduced to the touch screen environment via a pre-training program. The goal was to train subjects to touch the screen appropriately, to become familiar with the registration and feedback tones following a response, and to cope with one object and then two objects serially. The stimuli were patterns of dots.

Training. (a) First only one stimulus was presented and the subject was required to touch the stimulus for six successive trials. When the subject touched the stimulus, a registration tone sounded, followed by a feedback tone indicating a correct response (in the case of one item there are no degrees of freedom, and therefore, a correct response is the only response a subject can make). There was a five second interval between trials. A warning tone indicated that a new trial was about to begin. The position of the stimulus changed randomly after each trial. After subjects made six successive correct responses, they advanced to the next stage. At the bottom of the screen there was a horizontal white line which increased as subjects approached each criterion and decreased after error. We referred to this as the "analog display." This allowed subjects to keep track of their cumulative performance in the session. (b) A second stimulus was added (according to the colour sequence). The subject was required to report the

items in the correct order (first A then B for the forward condition, first D then E for the backward condition). Whenever a subject touched a stimulus, a registration tone sounded. In the case of an incorrect response, the registration tone was immediately followed by a feedback tone, the screen went blank, and the trial was repeated after the usual five second intertrial interval. If four successive errors were recorded, the subject returned to the previous stage (one stimulus). Six successive correct responses were required to return the subject to two stimuli. Correct feedback tones were only given at the completion of a fully correct string. To prevent any cues from spatial position, each element changes location after each correct trial. Again a criterion of six successive correct responses were required for advancement to the next stage. (c) A third stimulus was added. The subject proceeded as above. If four successive errors were recorded, the subject returned to the previous stage (two stimuli). Six successive correct responses were required to return the subject to three stimuli. In turn, six successive correct responses were required for advancement to the next stage. (d) As before, with four stimuli. (e) Finally the subject was confronted with the entire set of five stimuli. After subjects reached criterion with the five-item set they followed the verbal probe and over-training procedures (below).

Each subjects was given one session per day. Sessions were given on consecutive days (school days only) as far as was possible. Sessions were not to exceed 15 minutes in length and a session was terminated if subjects showed signs of tiring. A typical session consisted of 32 trials (excluding correction trials).

Verbal Probe. Verbal probe questions were asked (in the same session) immediately after subjects reached the training criterion with the full five-item (ABCDE) set. Subjects were turned away from the screen so that it was no longer visible. They were then asked the following questions.

A. (Free Recall) Can you tell me which colours you touched on the screen? (The experimenter pauses to allow the subject to freely recall the colours touched on the screen).

B. (Prompted) Which colour do you touch first? (pause) and then? (pause) and then? (pause) and then? (pause) and then? (pause).

C. (Prompted) Which colour do you touch last? (pause) and before that? (pause) and before that? (pause) and before that? (pause) and before that? (pause).

D. (Gap Filling) Imagine/Suppose/If you had only two colours on the screen, and one was D (the experimenter fills in the appropriate colour for the subject) and the other was B. Which one would you touch first? (pause) Which one would you touch last? (pause)

E. (Exclusion) If B and D are on the screen, which colours are left out/not there/missing?

Subject responses to these questions were recorded both on audio tape and protocol sheets ticked off in situ. No differential feedback was given, however subjects were given neutral praise and encouragement. Completion of the last question was followed by an invitation to "play a new game tomorrow." The session was terminated and the subject returned to the classroom.

Over-training. In order to check the reliability of performance, once subjects reached criterion with the entire set of five stimuli, in a separate session they were confronted with the entire set of five stimuli from the start. Subjects were required to restate the acquisition criterion of six successive correct trials. Subjects were tested further to see if they could meet a more stringent criterion of nine correct trials out of ten. After reaching criterion, subjects went on to test phase I in a further session.

Test Phase I. Fifteen test pair, triplet and quadruplet trials were derived from the total permutative set. These test trials were chosen because of their particular significance with respect to other studies of serial order in the comparative literature. A subset rather than the full set was used in order to obtain the maximum amount of information in the test phase without tiring young subjects. Test trials were presented in a session along with routine ABCDE training trials. The ratio of training to test trials was 50/50. A test session began with "warm-up" ABCDE training trials (no more than five). No direct differential feedback was given on test trials. However, indirect feedback was given in that the analog display recorded

correct performance only. On embedded training trials, the usual feedback tones were given. Correct performance on embedded training trials also affected the analog display. After completing test phase I, subjects went on to test phase II in further sessions.

Test Phase II. This phase was conducted in four parts, each in separate sessions.

Delayed Verbal Response. In the first part, subjects were confronted with what appeared to be a normal five-item (ABCDE) sequence. Whilst the subject was responding however, the screen was cleared at quasi-random points in the series, i.e. sometimes after a response to the first item, sometimes after responses to the first two items, the first three, etc. After each trial subjects was asked by the experimenter to name the colours which had not yet been touched.

Delayed Non-Verbal Response (no position changes). In the second part, subjects again began with what appeared to be a normal five-item (ABCDE) sequence and were interrupted at quasi-random points in the series. However, this time the items were re-presented after a delay of five seconds, during which a "bridging tone" was sounded. Upon re-presentation, the colours of the series remained in the same spatial locations as in the "priming" sequence. When the items were re-presented, the subject was required to "tell the machine" (by touching) which colours had not been touched on the first presentation. Subjects were allowed to report items in any order until they indicated that they were satisfied that they had completed the sequence. The subject terminated the trial by pressing the space bar. No differential feedback was given, however if subjects reported the relevant items in the correct order, the analog display was incremented affording the subject with indirect evidence of relative success.

Delayed Non-Verbal Response (with position changes). The third part followed the same procedure as the second, except upon re-presentation, the colours of the series assumed different spatial locations from the "priming" sequence.

Delayed Verbal Response. In the fourth part, subjects repeated the verbal version of the task as in part one.

Results

Acquisition

Subjects. Of the 12 subjects who began the study, one subject (aged 3,10) was dropped in the pre-training stage because he was unable to touch the screen appropriately. Of the 11 subjects who began training, all except one were able to reach the criterion with 5 items (ABCDE).¹

Trials to criterion. Total numbers of trials to the ABCDE criterion for each subject are shown in Table 4.2. Mean number of sessions required to learn the ABCDE series for the total group was 4. However, the forward training group required far fewer sessions than the backward training group. The mean number of sessions required by the forward training group was 1, whereas the mean for the backward training group was 6. The mean number of trials required to learn the ABCDE series for the total group was 43 (the minimum number of trials required to reach criterion was 24). The mean number of trials (excluding correction trials) required by the forward training group was 32, whereas the mean for the backward training group was 54. Subjects in the backward training condition required nearly twice (1.7) the number of trials to reach all the criteria than subjects in the forward condition.

¹Subjects Barry and Ben served as "prototype" subjects in the early phases of data collection. For this reason, some of the more detailed aspects of their data are unavailable.

			TRIALS TO	TRIALS TO	TRIALS TO	TRIALS TO	TOTAL TO
SUBJECT	AGE	COND	AB/DE	ABC/CDE	ABCD/ BCDE	ABCDE	ABCDE
BARRY	4,7	forward	N/A	N/A	N/A	N/A	N/A
JANE	4,6	forward	6	6	7	11	30
BEN	4,3	forward	5	5	9	6	25
LOUISE	4,1	forward	6	6	12	11	35
LEWIS	3,10	forward	6	6	19	8	39
KIRSTY	3,9	forward	6	10	9	8	33
PAULA	4,6	backw'd	6	8	6	15	35
CHRISTINA	4,6	backw'd	5	38	7	6	56
FINLAY	4,6	backw'd	36	10	6	N/A	52
AMY	3,11	backw'd	6	11	8	29	54
ANNA	3,10	backw'd	49	7	6	10	72
TOTALS		forward	29	33	56	44	162
		backw'd	102	74	33	60	269
		total	131	107	89	104	431
MEANS		forward	6	7	11	9	32
		backw'd	17	16	7	15	54
		total	13	11	9	12	43

Table 4.2: Trials to Criterion: Training

Breakdown of training criteria. Because the series was acquired in a parsed fashion, it was possible to examine numbers of trials required to reach criterion for each part of the series. For forward training group subjects these were the AB, ABC, ABCD, and ABCDE criteria and for backward training group subjects these were the DE, CDE, BCDE, and ABCDE criteria. In this way we can locate resistance points in acquiring the series both for groups and individuals. Table 4.3 shows the percent of total trials to criterion required to reach each of the subseries criteria both for groups and for individuals. A majority of the forward training group subjects had the greatest difficulty with the ABCD (4-item) series, in terms of a rise in the number of trials required to reach criterion. In fact, the largest percentage of training trials for the forward training group (35%) were spent acquiring the ABCD series, followed by the ABCDE series at 27%.

Acquisition of the AB and ABC series required 18% and 20% of the trials respectively. Expressed in terms of ranks (1 = least trials, 4 = most trials), AB = 1, ABC = 2 ABCD = 4, and ABCDE = 3. Analysis using Kendall's coefficient of concordance ($W = 0.63$) showed that this level of concordance was significant at the $p < 0.05$ level. Backward training group subject profiles were more varied ($W = 0.33$), however each subject had one subseries in particular which required considerably more trials than the others. Subjects Amy and Paula required the most trials to acquire the ABCDE series. In Amy's case 54% of her total trials to criterion were spent on ABCDE. For two subjects (Anna and Finlay), acquiring the DE (2-item) series required more than 3 times as many trials as any other series. This amounted to 68% and 69% (respectively) of all trials to criterion. Christina had particular trouble with the CDE series which required 4 times more trials (68%) to reach the criterion than any other series.

			TRIALS TO	TRIALS TO	TRIALS TO	TRIALS TO
SUBJECT	AGE	COND	AB/DE	ABC/CDE	ABCD/BCDE	ABCDE
BARRY	4,7	forward	N/A	N/A	N/A	N/A
JANE	4,6	forward	20%	20%	23%	37%
BEN	4,3	forward	20%	20%	36%	24%
LOUISE	4,1	forward	17%	17%	34%	31%
LEWIS	3,10	forward	15%	15%	49%	21%
KIRSTY	3,9	forward	18%	30%	27%	24%
PAULA	4,6	backward	17%	23%	17%	43%
CHRISTINA	4,6	backward	9%	68%	13%	11%
FINLAY	4,6	backward	69%	19%	12%	N/A
AMY	3,11	backward	11%	20%	15%	54%
ANNA	3,10	backward	68%	10%	8%	14%
TOTALS		forw'd tr	18%	20%	35%	27%
		back'd tr	38%	28%	12%	22%
		total	30%	25%	21%	24%
MEANS		forw'd tr	18%	20%	35%	27%
		back'd tr	31%	30%	13%	28%
		total	30%	25%	21%	27%

Table 4.3: Trials to Criterion: Distribution

Errors. Errors were divided into two types: forward errors and backward errors. Forward errors are errors in which the subject skips ahead in the series, e.g. ABD. Backward errors are errors in which the subject “backtracks” to items already reported, e.g. ABA. Analysis of errors accumulated over sessions of training on ABCDE trials revealed that 92% of errors committed by all subjects were of the forward type. Breakdown for the two training conditions showed that all errors made by the forward training group were of the forward type. All of these errors, except one, were of the one-step variety. For the backward training group however, 90% of errors were of the forward type and 10% were of the backward type. Further breakdown revealed that 31% of the forward errors were, in particular, errors in which the subject started to report the series in the wrong place. We called this particular type of error a *reference error*. Reference errors made of 28% of all the backward training group's errors. Forward training subjects made no errors of this type. Breakdown of the remaining forward type errors for the backward training group showed that 50% were of the one-step variety, 33% of the two-step variety, and 17% of the three-step variety.

Verbal Probes

Subjects. The 10 subjects who reached the ABCDE criterion were asked questions about the series. Results can be found in Table 4.4. Correct responses are shown in bold face type. Only one subject, Jane, did not participate due to an unwillingness to talk. Barry and Ben did not receive the full set of questions due to their status as “prototype” subjects.

SUBJECT	AGE	COND	FREE (ABCDE)	PROMPT (ABCDE)	PROMPT (EDCBA)	GAP (BD)	EXCL'N (ACE)
BARRY	4,7	forward	ABCDE	ABCDE	N/A	N/A	N/A
PAULA	4,6	backw'd	ABCDE	EDCBA	ABCDE	DB	ACE
CHRISTINA	4,6	backw'd	ABCDE	ABCDE	EAB	DB	ACE
BEN	4,3	forward	EDCBA	ABCED	A	N/A	N/A
LOUISE	4,1	forward	ABCDE	ABCDE	EDCBA	AE	AE
AMY	3,11	backw'd	ABCDE	ABCDE	EDABC	DB	AEC
LEWIS	3,10	forward	ABCDE	ABCDE	EA	BD	ACE
ANNA	3,10	backw'd	A	CDECB	CBACD	AC	D
KIRSTY	3,9	forward	ACBED	ABCDE	EDCBA	BD	CEA

Table 4.4: Verbal Probes

Free Classification. All subjects except one, Anna, were able to freely name all colours they had touched (she reported A only). She was the only subject unable to answer any of the questions correctly. The majority of subjects (6 of 9) named the colours in the correct ABCDE order, however two subjects, both from the forward training group, did not. Ben named the colours in the reverse order (EDCBA) and Kirsty reported the series as ACBED.

Prompted. All subjects who correctly reported the colours they had touched in order were also able to report them correctly when prompted for the first one, the next one, etc., except Paula who reported them in the reverse order (EDCBA). She was the only backward training group subject to show a predisposition to name colours in the order they were presented in training (as opposed to their order in the series). Only one subject, Kirsty who was unable to name the colours unprompted was able to do so when prompted. Ben reported ABC in the correct order, but D and E were incorrect. Only two subjects were able to correctly name the colours in the reverse order when prompted. Three further subjects reported E as last but were unable to correctly place all the items.

BD gap-skipping and exclusion. Only two subjects were able to correctly name B as first and D as last. Three other subjects reported the items in the order they were mentioned (DB). Louise reported that "A is always first" and "E is always last." Five subjects were able to correctly report the

correct colours A, C, and E were missing if only B and D were on the screen, with three subjects reporting them in order as ACE and two others reporting them as AEC and CEA. Louise reported A and E only.

Over-training (Routinization)

Subjects/Performance. All subjects who acquired the ABCDE criterion (the original eleven subjects less Finlay) were given at least one session of over-training trials. In the over-training trials, all subjects except one were able to restate the acquisition criterion and all but three were able to meet the more stringent nine correct trials out of ten over-training criterion.

Decision Time. Response time (RT) data were taken from the nine correct trials which made up the 9 out of 10 over-training criterion. In cases where this data was not available, the data were taken from 9 correct trials within the over-training phase. The mean response latency to each item of the series was calculated for each subject. Values greater than or equal to 3 times the standard deviation were replaced by the mean.

Average decision times for completion of the entire ABCDE series for can be found in Table 4.5. The average decision time was 10.38 seconds, with 11.48 seconds for the forward training group and 10.11 seconds for the backward training group. There was considerable difference in the time required for individual subjects to complete the series. The subject with the shortest decision time, Jane (6.19 seconds), was 7.71 seconds faster than the slowest subject, Barry (13.89 seconds). Both were forward training group subjects. There seemed to be no effect of age in decision time; both the slowest and the fastest subject were among the oldest in the study. Jane's rapid decision time was exceptional, especially when compared to other forward training group subjects. The next shortest decision time to hers was 11.25 seconds, a difference of 5.07 seconds. If Jane's decision time is discounted, there remains a difference of 2.64 seconds between the fastest and slowest forward training group subjects. In the backward training group there was a difference of 3.21 seconds between the slowest and fastest decision times.

Comparing the decision time of children with those found for monkeys & pigeons, it was found that children's decision times were slower when compared to those of *Cebus apella*, which averaged 3.75 seconds to

complete the ABCDE series (D'Amato & Colombo 1988, Figure 4.1). Also children were, on average, slower than pigeons which averaged 7.03 seconds to complete the ABCDE series.

Forward Training

Subject	A	B	C	D	E	Total
Barry	377 (27)	301 (22)	213 (15)	225 (16)	272 (20)	1389
Jane	135 (22)	132 (21)	140 (23)	107 (17)	105 (17)	619
Ben	403 (30)	228 (17)	277 (21)	286 (21)	147 (11)	1340
Louise	316 (27)	188 (16)	301 (26)	217 (19)	134 (12)	1156
Lewis	350 (28)	282 (22)	242 (19)	178 (14)	209 (17)	1260
Kirsty	276 (25)	185 (16)	244 (22)	247 (22)	173 (15)	1125
MEAN	309 (27)	219 (19)	236 (21)	210 (18)	173 (15)	1148

Backward Training

Subject	A	B	C	D	E	Total
Paula	296 (34)	145 (17)	114 (13)	183 (21)	126 (15)	864
Christina	268 (26)	258 (25)	172 (17)	191 (19)	134 (13)	1023
Amy	214 (22)	227 (23)	185 (19)	185 (19)	158 (16)	970
Anna	310 (26)	269 (23)	195 (16)	245 (21)	167 (14)	1185
MEAN	272 (27)	225 (22)	167 (16)	201 (20)	146 (15)	1011

Total	A	B	C	D	E	Total
MEAN	296 (27)	221 (20)	211 (19)	207 (19)	163 (15)	1038

Table 4.5: Average Inter-item RT expressed in hundredths of seconds (percent of decision time in parenthesis)

Inter-item RT. Table 4.5 shows the average inter-item RT for all subjects. For the forward training group average inter-item RT was 3.09, 2.19, 2.36, 2.10, and 1.73 seconds respectively and for the backward training group 2.72, 2.25, 1.67, 2.01, and 1.46 seconds respectively. The overall average inter-item RT was 2.25 seconds. The pattern of these latencies shows a pattern of self-imposed phrasing on the part of subjects. For the forward training group this pattern was AB-CDE and for the backward training group, ABC-DE (see Figure 4.1). However, long latencies to item A may

have been artificially inflated due in part to the fact that the onset of the stimulus array was not response contingent. Figures in parenthesis (Table 4.5) show inter-item RT expressed as a percent of decision time. All subjects (except Amy) spent the greatest percentage of total decision time in responding to A, an average of 27%. If the "expense" of inter-item RT is expressed in terms of ranks (1 = least time, 5 = most time), it was found that for the forward training group A = 5, B = 3, C = 4, D = 2, and E = 1, for the backward training group A = 5, B = 4, C = 2, D = 3, and E = 1, and for the entire group A = 5, B = 4, C = 2, D = 3, and E = 1. Analysis using Kendall's coefficient of concordance showed that the rankings of the both the forward training group ($W = 0.58$) and the backward training group ($W = 0.88$) were significant at the $p < 0.01$ level and the ranking of the entire group ($W = 0.61$) was significant at the $p < 0.001$ level.

Comparing the inter-item RT of children with those found for monkeys & pigeons, it was found that children's inter-item RTs were much slower those of *Cebus apella*. For the ABCDE series D'Amato & Colombo (1988) found that monkeys showed average inter-item RT of 1.50, 0.75, 0.50, 0.50, and 0.50 seconds respectively. Forty percent of total decision time was spent on the first item with roughly equal proportions distributed over the remaining elements. For the ABCDE series Terrace (1991, see Figure 4.2) found that pigeons showed average inter-item RT of 1.60, 1.30, 1.50, 1.50, and 1.20 seconds respectively. Pigeons did not show the large RT to A seen in monkeys and children, however, as with monkeys total decision time was evenly distributed.

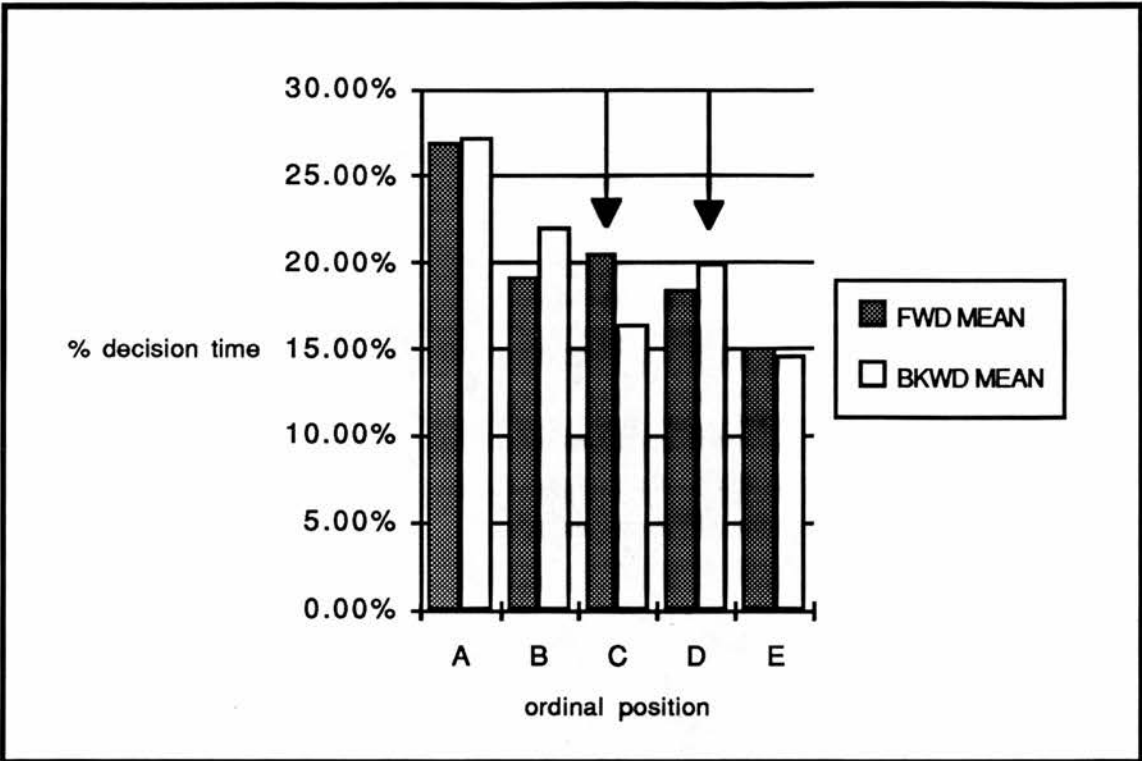


Figure 4.1: Inter-item RT as a percent of decision time: Forward v. Backward training

Test Phase I

Subjects. All subjects who acquired the five-item string ABCDE were given fifteen pair-wise, triplet, and quadruplet tests which were selected from the total permutative set.²

Performance. Table 4.6 shows the percent correct obtained for both test trials and interspersed training trials for each subject. All subjects showed a decrement in performance on test trials. With respect to the performance of forward training versus backward training group subjects, although their performance on interspersed training trials was roughly the same (86% and 87% correct respectively), the forward training group showed performance on test trials worse than that of the backward training group (55% and 68% correct respectively). However, a closer look at the data revealed that these decrements were more the result of poor

² Barry (and to a certain extent Ben) was a prototype subject and did not receive the same set of test trials as the other subjects. His data are not included with those of the other subjects.

performance by particular individuals rather than poor performance by the group as a whole. Three subjects performed particularly poorly (less than 20% correct) on the test trials. Two of these were from the forward training group and one was from the backward training group. Poor performance appears not to be the result of bad training performance. Only one of the three poor test performers also showed a poor performance on interspersed training trials (Kirsty 63% correct). Both of the others maintained better than 80% accuracy on interspersed training trials. Of the remaining six subjects, five were 70% correct or better on test trials. Two of these were from the forward training group and three were from the backward training group. One remaining forward training subjects scored 57% correct on test trials. All of these subjects were better than 80% correct on interspersed training trials.

Subjects	TEST	TRAINING	DECREMENT
jane	57%	80%	23%
lewis	79%	92%	14%
louise	71%	100%	29%
kirsty	0%	63%	63%
ben	18%	86%	68%
amy	71%	93%	22%
anna	17%	83%	67%
chris	85%	88%	4%
paula	71%	81%	10%
<i>forward</i>	<i>55%</i>	<i>86%</i>	<i>31%</i>
<i>backward</i>	<i>68%</i>	<i>87%</i>	<i>19%</i>
<i>total</i>	<i>61%</i>	<i>86%</i>	<i>25%</i>

Table 4.6: Test Performance by Subject

Further breakdown. Test performance was further analysed in three ways. Firstly, performance was broken down according to the number of constituents the subject was presented with, i.e. pair-wise, triplet, and quadruplet trials. Secondly performance was broken down according to the number of "gaps" between items, e.g. ABC is a zero-gap separation, ABD is a one-gap separation, ABE is a two gap separation. Finally, performance was broken down according to which item the subject was

meant to touch first, e.g. A-start, B-start, or C-start trials. The three subjects who showed very poor performance (less than 20% correct) on test trials showed across the board decrements regardless of trial type. Two of these subjects were the youngest of the group. They were removed from further breakdown percentages. Six subjects remained in the sample, three from the forward training group and three from the backward training group.

Table 4.7 shows the breakdown of performance for quadruplets, triplets, and pairs. Our subjects were 90% correct on quadruplet trials, 67% correct on triplet trials, and 78% correct on pair-wise trials. Although performance on triplets was worse than for other test trials, it must be remembered that there were more triplet trials than pairs and quadruplets.

Test Performance X string length			
Subjects	QUADS	TRIPLETS	PAIRS
jane	100%	40%	100%
lewis	100%	78%	67%
louise	50%	67%	100%
paula	100%	80%	33%
chris	100%	88%	67%
amy	100%	56%	100%
forward	80%	61%	89%
backw'd	100%	74%	67%
total	90%	67%	78%

Table 4.7: Performance on quadruplets, triplets, and pairs.

Table 4.8 shows the breakdown of performance for zero-gap, one-gap, and two-gap trials. Our subjects were 78% correct on zero-gap trials, 61% correct on one-gap trials, and 78% correct on two-gap trials. Performance on zero- and one-gap trials was similar across subjects, however on two-gap trials two subjects performed particularly poorly whilst the others were 100% correct.

Test Performance X "gaps" to be skipped			
Subjects	0	1	2
jane	83%	60%	0%
lewis	83%	60%	100%
louise	71%	50%	100%
paula	67%	60%	100%
chris	80%	80%	100%
amy	86%	50%	67%
forward	79 %	57 %	67 %
backw'd	78 %	64 %	89 %
total	78 %	61 %	78 %

Table 4.8: Breakdown of performance by "gaps"

Table 4.9 shows the breakdown of performance for A-start, B-start, and C-start trials. Our subjects were 80% correct on A-start trials, 68% correct on B-start trials, and 55% correct on C-start trials. It must be remembered that there were more test trials which began with A, rather than with B or C. However it is interesting that, in general, subjects performed best on test trials which contain A.

Test Performance X starting point			
Subjects	A	B	C
jane	57%	60%	50%
lewis	86%	60%	100%
louise	83%	67%	50%
paula	100%	60%	0%
chris	100%	80%	0%
amy	57%	80%	100%
forward	75 %	63 %	67 %
backw'd	86 %	73 %	40 %
total	80 %	68 %	55 %

Table 4.9 : Breakdown of performance on A-start, B-start, and C-start trials

Comparison with D'Amato & Colombo (1988) results and Terrace (1987; 1991). Table 4.10 summarizes our results along with those of D'Amato & Colombo (1988) and Terrace (1987; 1991). In terms of overall test performance, monkeys showed the highest percentage correct (87%) as compared with 72% correct for children and 58% for pigeons. It must be

remembered that D'Amato & Colombo (1988) and Terrace (1987; 1991) used the entire permutative set of test pairs and triplets, whilst we used a subset of the pairs, and added two quadruplets. Direct comparison was therefore limited. However, if only test trials common to all three studies are taken into consideration (interior pairs and all triplets, see Table 4.11), test performance for monkeys was 84% correct as compared with 70% for children and 44% for pigeons. This performance must also be viewed in the context of training performance. Both pigeons and monkeys were removed from ABCDE training to testing when they reached the 75% correct criterion, whereas the children in this study continued to receive training trials interspersed between test trials. On such interspersed test trials they were, as a group, 88% correct with no individual subject less than 80% correct. Whilst both children and pigeons showed a decrement in test performance compared to training performance (16% and 17% respectively), monkeys showed an increment (12%). If only trials common to all three studies are taken into consideration, the decrement is 18% for children and 31% for pigeons, with monkeys still showing an increment of 9%.

Test Performance			
Subjects	TEST	TRAINING	DECREMENT
children	72%	88%	16%
monkeys	87%	75%	-12%
pigeons	58%	75%	17%

Test Performance X string length			
Subjects	QUADS	TRIPLETS	PAIRS
children	90%	67%	78%
monkeys	###	85%	89%
pigeons	###	42%	74%

Test Performance X "gaps" to be skipped				
Subjects	0	1	2	3
children	78%	61%	78%	###
monkeys	80%	89%	92%	100%
pigeons	52%	51%	71%	90%

Test Performance X starting point				
Subjects	A	B	C	D
children	80%	68%	55%	###
monkeys	93%	88%	75%	65%
pigeons	70%	41%	47%	82%

Table 4.10: Comparison to Terrace and D'Amato & Colombo

Reported test performance for pigeons and monkeys was further analysed in three ways (as above) --- according to the number of constituents, the number of "gaps" between items, and according to first item. With respect to performance according to the number of constituents, all three groups performed better on pair-wise tests than triplets. However the improvement in performance varied from 4% better for monkeys to 11% for children and 31% for pigeons. If only common trials are taken into

consideration, the gap between pair and triplet performance narrows for both pigeons and monkeys. Pigeons show 47% correct for pairs and 42% correct for triplets, whilst monkeys show 82% correct for pairs and 85% correct for triplets. With respect to performance according to the number of "gaps" between items, both pigeons and monkeys show an increase in percent correct along with an increase in the number of gaps, whilst children show a decrease for "one-gap" trials. If only common trials are taken into consideration, the monkeys maintain a small increase (79%, 87%, 89%) whilst the pigeons show a marked increase in percent correct for "two-gap" trials (39%, 36%, 63%). Finally, with respect to first item, both children and monkeys were best at test trials which started with A and showed a steady decrease on test trials starting with B and C (and D for monkeys). Pigeons on the other hand showed their best performance on test trials starting with D (82%), second best on test trials starting with A (70%) and worst on trials starting with B and C (41% and 47% respectively). If only common trials are taken into consideration, all three groups show their best performance on trials starting with A, next best on trials starting with B, and worst on trials starting with C.

Test Performance			
Subjects	TEST	TRAINING	DECREMENT
children	70%	88%	18%
monkeys	84%	75%	- 9 %
pigeons	44%	75%	31%

Test Performance X string length		
Subjects	TRIPLETS	PAIRS
children	67%	78%
monkeys	85%	82%
pigeons	42%	47%

Test Performance X "gaps" to be skipped			
Subjects	0	1	2
children	74%	61%	78%
monkeys	79%	87%	89%
pigeons	39%	36%	63%

Test Performance X starting point			
Subjects	A	B	C
children	77%	67%	55%
monkeys	88%	87%	68%
pigeons	57%	33%	31%

Table 4.11: Comparison to Terrace and D'Amato & Colombo (common trials only)

Test Phase II

Subjects. All subjects who completed the first test phase participated in the second test phase, with the following exceptions. Subjects Anna and Ben left the study at this point due to a lack of cooperation on their part. Also Jane did not participate due to an unwillingness to speak.

Performance. Table 4.12 shows percent correct obtained for both verbal and non-verbal test trials. The requirements were that the subject had to report the remainder of the series in the correct (ABCDE) order. The overall percentage correct was 53% correct with 53% correct for the forward training group and 52% correct for the backward training group. The breakdown for verbal trials was 56% correct for the forward training group and 53% correct for the backward training group. Scores for both sessions of verbal trials can be found in Table 4.12 (bottom panel). The breakdown for non-verbal trials was 50% correct for training groups. Breakdown of scores for non-verbal trials, both with and without position changes, can be found in Table 4.12 (bottom panel). Position changes resulted in a 19% decrement in performance for the forward training group, but a 15% improvement in performance for the backward training group. Although there was little variance between performance on verbal and non-verbal tasks among groups, one subject in particular, Lewis, showed high levels of performance on verbal conditions (90% and 100%), but low levels of performance on non-verbal conditions (25% and 22%).

Total Group	Verbal	Non Verbal	All Trials
trials	140	118	258
errors	63	59	122
%correct	55%	50%	53%
Forward Tr	Verbal	Non Verbal	All Trials
trials	80	64	144
errors	35	32	67
%correct	56%	50%	53%
Backw'd Tr	Verbal	Non Verbal	All Trials
trials	60	54	114
errors	28	27	55
%correct	53%	50%	52%

Total Group	Verbal 1	NonVerbal N	NonVerbal P	Verbal 2	All Trials
trials	70	60	58	70	258
errors	33	29	30	30	122
%correct	53%	52%	48%	57%	53%
Forward Tr	Verbal 1	NonVerbal N	NonVerbal P	Verbal 2	All Trials
trials	40	30	34	40	144
errors	17	12	20	18	67
%correct	58%	60%	41%	55%	53%
Backw'd Tr	Verbal 1	NonVerbal N	NonVerbal P	Verbal 2	All Trials
trials	30	30	24	30	114
errors	16	17	10	12	55
%correct	47%	43%	58%	60%	52%

Table 4.12: Test Phase 2 Summary

Error and ordinal position. What was the effect of the ordinal position of the delay on the ability of the subject to complete the sequence? For one subject, Kirsty, error was across the board. For this reason her data were removed from further breakdowns according to ordinal position. First we looked at the percentage of correct series completions at each delay position, and then we looked at how the error distributed across ordinal position. The distribution of error across ordinal position for both training groups is shown in Figure 4.2. For all trials, the subjects as a group were 41% correct when the delay was after A, 36% correct when the delay was after AB, 40% correct when the delay was after ABC, 73% correct when the delay was after ABCD, and 75% correct after ABCDE. Incorrect series completions were distributed 30% after A, 29% after AB, 28% after ABC, 6% after ABCD, 7% after ABCDE. The forward training group were 47% correct after A, 48% after AB, 41% correct after ABC, 59% after ABCD,

and 72% correct after ABCDE. Incorrect series completions were distributed 29% after A, 26% after AB, 32% after ABC, 10% after ABCD, 3% after ABCDE. Error after ABCDE was due to one error by one subject. Note that there was a large proportion of error after ABC. Recall that in over-training analysis of inter-item RT revealed a “phrase marker” at C. Having to start on D constituted breaking into a phrase. The backward training group were 35% correct after A, 19% correct after AB, 39% correct after ABC, 91% correct after ABCD, and 77% correct after ABCDE. Incorrect series completions were distributed 31% after A, 31% after AB, 25% after ABC, 4% after ABCD, 9% after ABCDE. Error after ABCD was due to one subject only. Note that there was a large proportion of error after AB. Analysis of inter-item RT revealed a “phrase marker” at D. Having to start at C constituted breaking into a phrase.

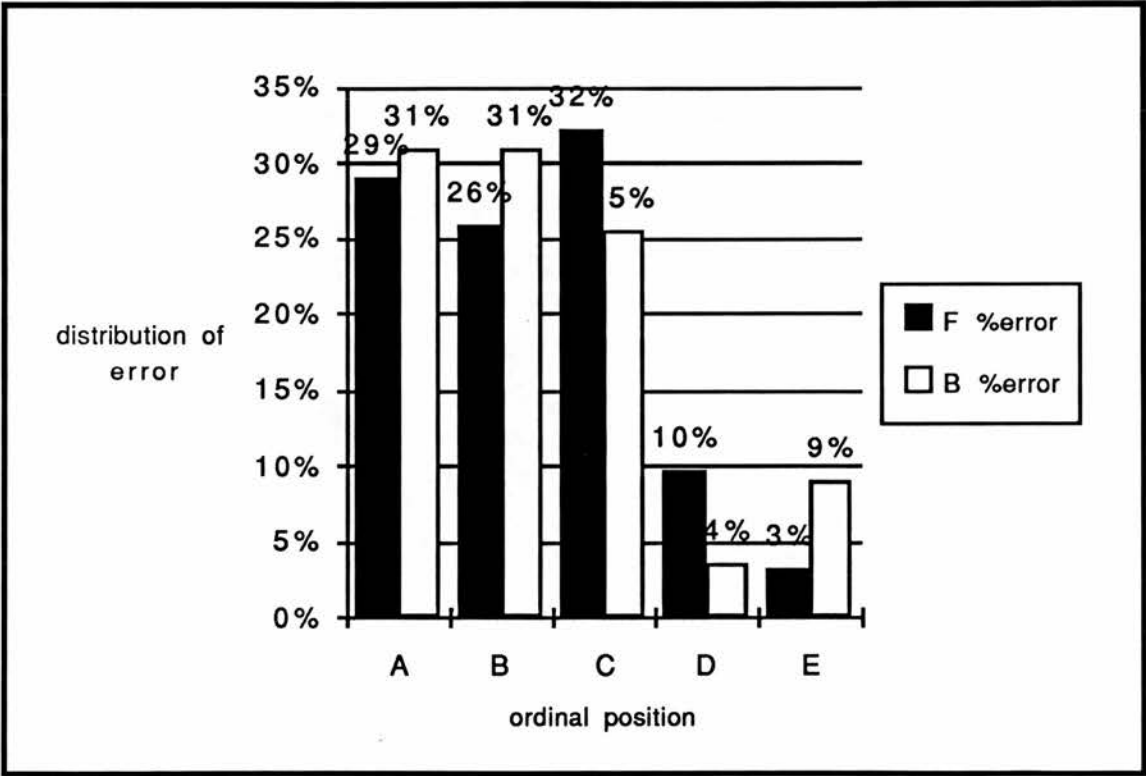


Figure 4.2: Series Completion: Distribution of total error

Next item correct. Note that for both groups the greatest proportion of error was concentrated on items A, B and C, indicating a strong effect of ordinal position. However, these items are at the beginning of the series. The dearth of error after D and E may have been due to the fact that there were fewer items left to report (one and none, respectively), fewer degrees

of freedom, and thus, less opportunity to err. However the degrees of freedom can be equalized by looking at the effect of the ordinal position of the delay on the ability of the subject to complete (at least) the *next* item in the sequence. Analysis revealed that only 6 errors fit this description (2 after A, 3 after AB, 1 after ABCD). Only the backward training group was affected. The overall percent correct becomes 45% correct after A, 42% correct after AB, 44% correct after ABC, 76% correct after ABCD, and 75% correct after ABCDE. This amounts to a 4% increase after A, a 6% increase after AB, a 4% increase after ABC, a 2% increase after ABCD. Incorrect series were distributed 30% after A, 28% after AB, 30% after ABC, 5% after ABCD, and 8% after ABCDE.

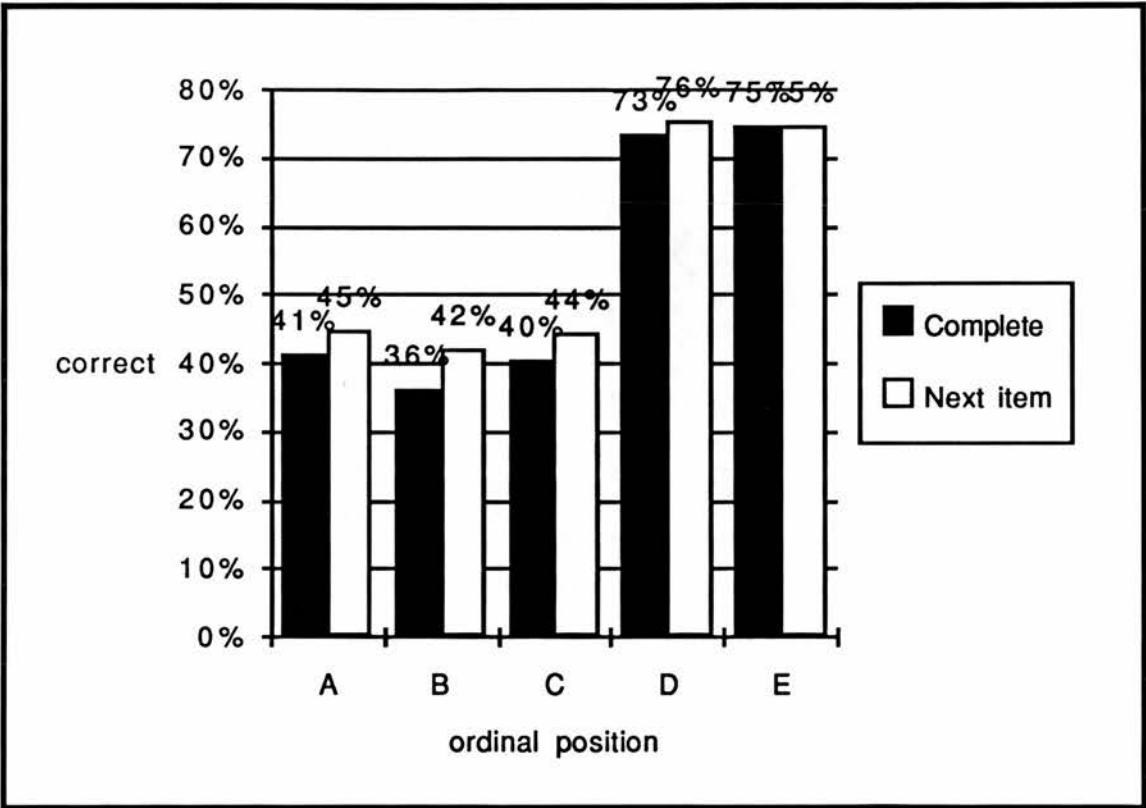


Figure 4.3: Series Completion v. Next item correct

Error types. Eight different types of error were observed and are presented in Table 4.13. The most frequent type of error observed was the “out of order” error in which the subject reports the remaining items, but not in ABCDE order. “Out of order” errors accounted for 38% of all errors. The next most frequent type of error observed was the “repeat” error in which the subject reports ABCDE regardless of the delay point. “Repeat” errors

accounted for 20% of all errors. However certain types of errors tended to be typical of individual subjects. All three backward training subjects tended to make "out of order" errors. The majority of Paula and Christina's errors were of the "out of order" type (59% and 57% respectively). Amy also made many "out of order" errors (31%). Forward training group error profiles were more subject to the individual. Kirsty, who was incorrect on all trials, typically reported only the last item she had touched before the delay. All of Lewis' errors on non-verbal trials were of the "repeat" type (he made only one "unclassified" error on the verbal trials). Louise, like the backward training subjects, made many "out of order" errors (50%). Barry made very few errors, however half were of the "hangover" type; he completed the series correctly, but added the last item before the delay to his response.

error type	forw'd	back'd	total
<i>out of order</i>	26%	45%	38%
<i>next</i>	0%	11%	7%
<i>hangover</i>	10%	7%	8%
<i>repeat</i>	35%	11%	20%
<i>"don't know"</i>	3%	4%	3%
<i>"none"</i>	3%	2%	2%
<i>unclassified</i>	19%	20%	20%
<i>omission</i>	3%	0%	1%

error type	barry	lewis	louise	paula	chris.	amy
<i>out of order</i>	0%	0%	50%	59%	57%	31%
<i>next</i>	0%	0%	0%	14%	0%	12%
<i>hangover</i>	50%	0%	6%	0%	14%	12%
<i>repeat</i>	0%	91%	6%	5%	29%	12%
<i>"don't know"</i>	25%	0%	0%	5%	0%	4%
<i>"none"</i>	0%	0%	6%	5%	0%	0%
<i>unclassified</i>	0%	9%	31%	14%	0%	31%
<i>omission</i>	25%	0%	0%	0%	0%	0%

Table 4.13: Error Type Summary

Error Types

1. Out of order. The subject reports the remaining items, but not in ABCDE order.
2. Next. The subject correctly reports one or more items remaining, but fails to complete the series correctly.
3. Hangover. The subject reports the item immediately prior to the delay.
4. Repeat. The subject reports ABCDE regardless of the delay point.
5. "Don't know." The subject claims not to know the correct answer.
6. "None." The subject wrongly claims that there are no items to report.
7. Omission. The subject misses out one or more items.
8. Unclassified errors.

Results Summary

4 year old children can learn a five-item (ABCDE) series using a non-verbal paradigm similar to that used by Terrace with pigeons (Straub & Terrace, 1981; Terrace, 1987; 1991) and D'Amato & Colombo (1988) with monkeys. Of the two training conditions employed, we found that subjects in the forward training condition required half as many trials as were required by backward training condition subjects to learn the five-item (ABCDE) series.

Failure to complete the ABCDE series was primarily due to the child's tendency to skip one item forward in the sequence. This was true in virtually every case for the forward training group. For backward training group subjects an additional source of error was the child's tendency to start reporting the series in the wrong place.

Subjects possessed limited ability to express knowledge about the ABCDE series verbally. Six out of nine subjects were able to verbally repeat the series both freely and when prompted for the items from first to last. However few subjects were able to report the series from last to first or perform correctly on "gap-skipping" and exclusion questions.

In over-training trials, seven out of 10 subjects were able to reach a nine correct trials out of ten over-training criterion. In these trials, subjects averaged 11 seconds to complete the ABCDE series, compared to 3.75 seconds for monkeys and 7.03 seconds for pigeons. Analysis of inter-item RT showed different phrasing effects for both the forward and backward training groups, with the forward training group phrasing the series as AB-CDE and the backward training group phrasing the series as ABC-DE.

In the first test phase, five of nine subjects achieved more than 70% accuracy on test trials (pairs, triplets, and quadruplets). Success or failure on test trials was not related to training condition. Failure on test trials (in all cases except perhaps one) was not the result of poor performance on interspersed training trials. However, success on training trials was no guarantee that a subject could perform well on the test trials. All other subjects, regardless of test success or failure, maintained better than 80% accuracy on interspersed training trials. Breakdown of test performance showed that subjects showed decrements in performance according to the

distance from A of the first item to report in the test trial. Examination of data from studies with pigeons and monkeys revealed this pattern of error was shown by monkeys, but not by pigeons.

In the second test phase, subjects were on average 53% correct on test trials (both verbal and non-verbal) in which the subject was required to complete the series after a delay. The greatest proportion of incorrect series completions were found when subjects were interrupted after A, AB, and ABC. For forward training subjects the greatest proportion of error was found when the delay was after ABC. For the backward training subjects the greatest proportion of error was found when the delay was after AB. In both cases interruptions at this point required the subject to resume reporting the series in the middle of a "phrase" as was found in analysis of inter-item RT in over-training.

Discussion

Training

Four year old subjects can learn to report a series of five unrelated elements. Using the same kind of non-verbal tests used in animal studies, it has been shown, that four year old children can learn (at least) a five-item (ABCDE) series to a reasonable level of accuracy. Moreover, subjects were successful using two different training conditions --- one in which the string was parsed forward from A and one in which the string was parsed backward from E. Subjects were trained using these two conditions in order to discover which training conditions yielded the best results in child subjects. Previously Straub et al. (1979) attempted to use the backward training condition with pigeons, but abandoned it after the pigeons' lack of success. Although our backward training condition subjects were "slower" to acquire the ABCDE series than forward training condition subjects in terms of numbers of sessions and trials to criterion, they were nevertheless able to acquire the series and maintain the same high level of performance as subjects trained in the forward training condition --- 80% or better accuracy through the test phase. Only one of our subjects (a backward training subject) was unable to reach criterion on the ABCDE series.

Verbal Protocols

If children are using some degree of verbal mediation to report the series, it does not give them an advantage, in terms of constituent numbers, over monkeys. Whilst monkeys and pigeons are obviously non-verbal subjects, there is the possibility that our child subjects may have used verbal mediation to report the series. Our results indicate that they had some verbal knowledge of the ABCDE series in that they could name the colours and verbally report them ~~the~~ the correct order. However, this knowledge was limited. Our subjects showed a clear prejudice toward starting at A and naming the colours in a "forward" direction. Our subjects could not verbally report the series backward. It would seem that for our subjects, such serial codes are represented unidirectionally and resist reversal³. Neither could our subjects correctly complete verbal "gap-skipping" or exclusion questions. Whatever the degree of verbal mediation used by the child subjects in this study, it did not give them an advantage over non-verbal primate subjects (*Cebus apella*) in terms of the number of constituents they could report in a series. However, there is a difference in constituent capacity between monkeys and children on the one hand (5 items plus), and pigeons on the other (4 to 5 items maximum). We interpret this as indicating that whatever difference exists in the representation of serial order between primates and pigeons, it is unlikely to be verbal in origin.

Reaction Time in Over-training

Our child subjects showed spontaneous phrasing in reporting the ABCDE series. There was a difference in the phrasing pattern of forward training subjects (AB-CDE) from that of backward training subjects (ABC-DE). However this "difference" may be the results of subjects dividing the constituents into the first two and then the next three. For the backward training subjects, D and E preceded A, B and C in training. Monkeys and pigeons do not show spontaneous phrasing. They tend to distribute effort in terms of inter-item RT roughly evenly after item A. Phrasing on the part of children may reflect a "natural rhythm," whereas non-human's

³ See Lashley's (1960) comments about the difficulty of playing a tune backwards.

even distribution of inter-item RT may be interpreted as reflecting that equal representative weight is tied to each item. This may be related to item capacity. Phrasing puts less strain on memory, whereas accessing each item individually would produce strain and ultimately limitations on the number of items subjects can report.

The speed with which monkeys (and to certain extent pigeons) complete the task together with the even distribution of inter-item RT may indicate an intolerance of delay; i.e. they move this quickly through the series because they have to. The natural pauses which are part of the rhythm of phrasing might be enough time to cause animals to lose their place in the series. This may explain the much faster decision time of monkeys and pigeons compared with our child subjects. However we have based our interpretation on the averaged group data found in the literature. We do not know if monkeys or pigeons would show phrasing effects if individual data were presented.

That monkeys and children show a larger proportion of total decision time spent on A (40% and 27% respectively) than pigeons do (22%) may indicate a species difference in strategy used to report the series. Although some of this time increase can be accounted for by stimulus onset, it is interesting that, although procedures were very similar for all three species, monkeys and children concentrate more "effort" on A than pigeons. This may indicate a strategy of reviewing the whole array before beginning to actually report the series.

Test Phase I

Good training performance is no guarantee of good test performance. Of the three subjects who performed poorly (less than 20% correct) on test trials only one of these also showed poor performance on interspersed training trials. The two other poor performers however were able to maintain performance at 83% and 86% correct on interspersed training trials. Therefore it is possible for a subject to maintain high performance on ABCDE trials whilst performing well below chance on test trials. Overall "headcounts" of how many individual subjects began training, how many completed it, how many of those entered the test phase and how many show overall test failure was not reported in comparative studies with pigeons or monkeys.

Test performance did not reveal differences between forward training and backward training subjects. Both poor and successful test performers were evenly distributed among the forward and backward training groups. Performance on individual test trials did not reveal wholesale differences between the two groups. Exceptions were the poor performance of backward training subjects on CD trials and the poor performance of forward training subjects on BCDE trials. Neither did further breakdown of test performance reveal differences between forward and backward training groups in terms of their ability to perform on test trials on the basis of the number of constituents, the number of "gaps," or the first item to be reported. The hypothesis that the ability to correctly "skip gaps" on test items is dependent on forward training was disconfirmed. Any apparent deficits between groups were consistently found to be the result of individual performance.

Starting position is an important factor for accuracy in test trials. Breakdown of performance on test trials indicates a steady decrease in accuracy on test trials in which the subject must start on an item other than A. This was true for both forward and backward training subjects. Although differences in number of constituents and the number of "gaps to be skipped" seemed to be contributory factors to failure, further analysis revealed that trials which started with items other than A were the source of these differences. High performance on trials beginning with A indicates the special status of A as an "anchor," even for subjects who were trained with the backward training programme (which forced them to shift "anchors" each time an additional item was added to the series). However it appears that once the full five-item set (ABCDE) was established, A acquired a special status. The fact that performance declined as subjects had to start deeper and deeper within the series indicates that as "distance" from A increases, the ability for the subject to use that item to "open" the series decreases.

Examination of the test errors of individual subjects indicates that A and E may have special status as the "entry" and "exit" elements of the series. A particularly striking example is Jane whose test errors reflect the special status of A and E vis a vis the interior items B, C, and D. Microanalysis of Jane's test errors revealed that she was 100% wrong on every test trial

which contained E. Her test errors were such that if A was present she touched it first and then touched E second. If A was not present she touched E first and then the others in the "correct" order (e.g. for BCE she chose EBC). On interspersed training trials, her errors were always incorrect choices to E. She, in particular, may have "represented" the ABCDE series as entry, exit, and interior. This means of reporting the series was stressed by the absence of some items in the test trials, resulting in a particular pattern of error. The pattern of test errors of other subjects also demonstrate that whilst A and E are firmly "represented," the interior items have an unsure status. Lewis' test errors typically involved swapping B and C, whilst Louise's test errors typically involved misplacement of C. Amy, unlike the other backward training subjects (and also unlike other subjects in general), showed better performance on test trials which started with C and worse performance on test trials which started with A. However microanalysis of errors on trials which started with A revealed that the errors were not due to incorrect touches to A, but rather to the other items. Paula typically made errors on test trials in which the first item was not A. In every case she chose D first.

Test Phase II

Delayed response tasks constituted a further exploration of the potential of the paradigm. In particular, we sought alternatives to the "gap-skipping" kinds of test trials used in the comparative literature --- a further test phase which could test subjects' knowledge of the ABCDE series without reducing the degrees of freedom. Overall performance was not very high, 53% correct overall. It is possible that the fixed delay of five seconds was too high, producing a floor effect in some subjects. A solution would be to have a variable delay, in which subject begin with short delays and gradually progress to longer delays, seeking their own levels of toleration. The delay after ABCDE was a bit of a "trick question." It was designed to find out if subjects knew where the end of the series was. In all other trials, the program terminated the series after the subject touched E.

Performance was selective according to the ordinal position of the delay. For both groups error was distributed almost evenly across A, B, and C (that is, delays which occurred after A, AB, and ABC), with very small amounts of error attributed to D and E. Error analysis showed that the

effect of ordinal position was stronger than the effect of degrees of freedom. However, further breakdown of error according to training condition showed that phrasing also had an effect. The places where subjects found it most difficult to correctly resume the series were places in which they were breaking into a phrase. For forward training subjects there was a large proportion of error after ABC. In the analysis of inter-item RT we found a "new phrase marker" at C. Having to start on D constituted breaking into a phrase (AB-CDE). For backward training subjects there was a large proportion of error after AB. In the analysis of inter-item RT we found a "new phrase marker" at D. Having to start at C constituted breaking into a phrase (ABC-DE).

The effect of ordinal position appeared to be stronger than that of phrasing. Subjects also showed poor performance in places where we would have expected better performance based on the fact that the subject would be beginning a new phrase. One would have expected that forward training subjects would show good performance in delays after AB since resuming on C would mean resuming on a new phrase. In fact subjects were only 48% correct in this position. Also, one would have expected that backward training subjects would show good performance in delays after ABC since resuming on D would mean resuming on a new phrase. In fact subjects were only 39% correct in this position. For all subjects performance was poor on trials in which delays occurred after A, AB, or ABC. For the forward training group the best performance of these three was after AB, and for the backward training group the best performance of these three was after ABC. This indicates that phrasing has some effect, but not greater than that of ordinal position.

Another example of the effect of ordinal position is the high level of performance found when the break was after ABCD. This is most likely due to the special status of E. Although subjects were required to break into a phrase, the special status of the last item overrides phrasing. This is especially true for backward training subjects, for whom E is especially salient. Therefore we see two effects conspiring to produce the pattern of error --- ordinal position and phrasing.

Error patterns of certain subjects may indicate strategies for dealing with delay. The large number of "out of order" errors made by backward

training subjects may indicate that for them the sequence is not necessarily stored as ABCDE in strict order. Consistent with their training, they typically fronted items E and D. The consistent "repeat errors" of Lewis may indicate that he cannot tolerate any break-up of the series, but must recall the series in its entirety. Although he showed few errors in the verbal delay conditions, review of his performance indicated that he may have been verbally rehearsing the entire set silently. "Hangover errors" made by some subjects may indicate that they are remembering the last item touched as a cue to start recalling from after the delay.

General Discussion

The touch screen technique works. Subjects responded well to the touch screen apparatus. They were keen to participate and maintained an interest throughout training and testing. It was possible to collect a large number of responses from individual subjects, much more than by conventional means. The results of using the touch screen technique in terms of data yield highlight it as one which is viable, and licenses its further use in future serial order and related investigations.

Conclusions about training direction. The forward training condition was not found to be responsible for certain test effects, nor was "gap skipping" found to be dependent upon forward training. However it was found that subjects required fewer trials to reach criterion when trained using the forward training programme. Considering that similar results were obtained after training, we concluded that because the forward training condition was more efficient and easier for subjects, it would yield the best results in future investigations using children as subjects.

Why was the backward training condition harder? In terms of the need for nearly double the trials to reach criterion, the backward training regime was much more difficult for our subjects than the forward training condition. Why? In order to answer this question we must examine what means subjects may have used to acquire the series.

For the forward group, failure to complete the ABCDE series was primarily due to children's tendency to skip one item forward in the sequence. Similar error type results were found by D'Amato & Colombo (1988) for monkeys and by Terrace (Straub & Terrace, 1981; Terrace, 1987; 1991) for

pigeons. During training on the ABC and ABCD series, forward errors were also the most prevalent. Furthermore, with respect to the ordinal position of errors, all our forward training subjects were 100% correct on touches to the first item (A) in ABCDE training and 96%-97% correct throughout AB, ABC, and ABCD training. During training, they virtually never made a reference type error. However, a substantial proportion (28%) of backward training subjects' errors on ABCDE training trials were reference type errors. This was not surprising since the first item changed every time the subject advanced to a larger number of stimuli.

We interpret these results as indicating that a strong directional element is involved in learning a series and that subjects seem to rely on the first item as an "anchor." Backward training subjects' main difficulty was that because the first item changed every time there was an increase in stimuli, they had no consistent "anchor" item from which to drive the series. This resulted in a larger number of errors than those subjects whose first item was unchanging. It is interesting to note that although the terminal item (E) was consistent for backward training subjects, it did not seem to serve the same "anchor" role as the first item did for forward training subjects. Results indicate however that by the time backward training subjects were training on the ABCDE series, two subjects were 100% correct in selecting A first (the other two were only 80% correct). From this it may be interpreted that these two subjects developed a strategy of always touching the novel item first.

Nevertheless, once the full five-item ABCDE set has been acquired differences seem to have diminished. Once there was a stable initial item, subjects were able to solidify their "representation." Forward and backward acquisition subjects had the same proportion of forward type errors and reference type errors (85% forward, 15% reference). The proportion of reference errors in the backward training group was reduced from 28% to 15%.

Comparison of test performance with that of pigeons and monkeys. It must be remembered that D'Amato & Colombo (1988) and Terrace (1987, 1991) used the entire permutative set of test pairs and triplets, whilst we used only the interior pairs and the entire set of triplets, plus two quadruplets. Whilst direct comparison is thus limited, there exists the

possibility for some interesting comparison between species. Our subjects' test performance must be seen in light of the fact that they were not tested on non-interior pairs. For both pigeons and monkeys performance on non-interior pairs was better than for interior pairs and tended to bolster up overall test performance. However when non-interior pair-wise trials were removed from monkey and pigeon test scores the resulting drop in overall percentage correct was only 3% for monkeys (from 87% correct to 84% correct) but 14% for pigeons (from 58% correct to 44% correct). Clearly the inclusion of non-interior pairs accounted for a large part of pigeons' success on test trials.

With respect to the success of monkeys compared to the children in this study it must be remembered that the monkeys are adult whilst four year old children are still in a developmental state. It would be interesting to compare the performance of human adults with those of monkeys on such serial order tasks. That monkeys performed better at test trials may indicate that monkeys were actually performing at better than the 75% correct criterion for ABCDE training. It would be interesting to see how monkeys would have performed on ABCDE training trials interspersed between test trials or on ABCDE training trials administered after the test sessions.

With respect to performance differences according to the first item of the test trial, both children and monkeys were best at test trials which started with A and showed a steady decrease on test trials starting with B and C (and D for monkeys). Pigeons on the other hand, showed their best performance on test trials starting with D, second best on test trials starting with A, and worst on trials starting with B and C. When only common trials are taken into consideration the special status of A is even more apparent. However, the success that pigeons had with trials beginning with end items (see performance on the DE pair) may indicate that E may have a special status as well as A for the pigeon. This is different from child and monkey evidence in which A has special status, and although the pattern of test errors in some children seems to indicate that E has a special status as the "exit" item, it is not likely to play the same role it does for the pigeon. This, along with the particularly poor performance of pigeons on interior pairs, may be interpreted as indicating that pigeons are using different means than monkeys and children to solve test trials.

"Chance performance" with respect to high levels of performance for the first item. Straub & Terrace (1981) treat AB, ABC, and ABCD as separate lists when calculating chance levels of performance. However due to the nature of the training paradigm, subjects have such considerable experience with AB before training on ABC, etc, that such subsequent lists should not be seen as separate events. Performance on ABC can only be evaluated in light of the fact that the subject has already had experience in responding to AB. Therefore on AB training trials the level of chance performance is 50% because there is only one degree of freedom, i.e. for the first response. Thereafter, if the subject correctly chooses A there is no way to make an error (since repeat touches are allowed to A). Consider ABC trials, where subjects by this time have considerable experience in touching item A first (in the forward training case). Because of this experience, one would expect very few reference errors (that is, errors at A). In fact, our forward training subjects were accurate on all touches to A (with one exception) in ABC training trials, effectively reducing the degrees of freedom from two to one and effectively maintaining chance odds at 50%.

Subjects might further improve their performance without particular knowledge about ABC as a series. In training on ABC trials our subjects virtually always respond to A and C correctly. Except for one error (a reference error to C) all other errors were of the AC type. We might interpret this to mean that subjects learnt something about entry and exit points. So long as subjects have a high rate of response to these items, they could achieve very high levels of accuracy on the ABC series. Difficulty in this strategy begins with the four-item ABCD series. However, if high response to A is maintained (e.g. accuracy for touches to A on ABCD trials for our subjects was 96%), the only real choice is between B and C, the interior items (D becomes the exit point). In this way subjects could, by prudent "gambling", achieve levels of accuracy above 50% without particular knowledge about ABCD as a series. If subjects were to develop an association between A and B even higher levels might be achieved, still without representing ABCD as a series. With five items however, this kind of strategy will come up against too many degrees of freedom to support high levels of performance without knowing something about ABCDE as a series.

Terrace (1991) admits that four- and five-item lists may tax the pigeon's memory span. He reports that only five of eight subjects were able to satisfy an accuracy criterion of 70% on ABCD series within 120 sessions. This level of accuracy seems close to what could be gained entirely through "gambling" strategies. One would predict from such a strategy that certain types of errors would be more likely than others on ABCD trials.

One would not expect many reference errors due to the high degree of routinization in choosing A first. One would expect many AC errors which are the result of gambling on C as the second choice rather than B. One would also expect forward errors to D, which can be interpreted as the subject skipping over the problematic interior items, but correctly selecting the "entry" and "exit" points (so to speak). Also one would predict poor performance in the test phase on interior pairs and, in the five item case, on the interior triplet BCD. In fact all three studies report that most errors are of the forward type, however the pigeon studies in particular feature poor test performance on interior pairs and on the BCD interior triplet. Without the support of the entry and exit items, the pigeon is reduced to chance performance. This can be seen in pigeon's test performance on interior pairs. In contrast, monkeys and children show good performance on interior pairs and the BCD triplet, which may be interpreted as indicating that something else is happening. Also it would be interesting to see what number of items would begin to produce performance deficits in monkeys and children as is seen for strings of four or five items with pigeons.

What are the implications of success on "gap-skipping" transfer trials? Firstly, success on test trials is not inevitable after success in learning the ABCDE series. We had two subjects who showed very poor performance on test trials despite being able to complete the ABCDE series. Neither Terrace nor D'Amato & Colombo report if there were any subjects who completed the training phase successfully, but showed poor test performance. They do not report a "drop out" rate. Neither do they report any individual data. It is possible that the test effects reported by Terrace and D'Amato & Colombo are a group artefact.

Secondly one must look at the results of the test phase *vis a vis* the results of the training phase. Both pigeons and children showed worse performance (in terms of percent of trials correct). Why did monkeys improve their performance? It is possible that the lesser number of items in test trials (two or three as opposed to the five items in an ABCDE trial) actually served to make test trials *easier* than training trials. It is not clear whether the subjects themselves make a connection between what is learned in the training phase and what is being tested by pair-wise and triplet test trials in the test phase. Consider subjects sophisticated at reporting lists of items. If these subjects were given test trials from a list not in their repertoire, how would their performance compare to subjects given test trials from a list already in their repertoire?

Also, we have already seen the questionable nature of success on trials containing A and/or E. It is possible that subjects could still do well on test trials whilst knowing only a some item information, e.g. the positions of A & C or A & E.

Considering the problems of reduced degrees of freedom and the degree to which subjects can use minimal knowledge of item position to produce seemingly good test profiles, the status of "gap-skipping" transfer trials as the ultimate test of "knowing" a series is dubious.

Alternatives to "gap-skipping" transfer tests. A heavier burden should be placed on performance on the full set. Ultimately, this is the best test of "knowing" a series. There is important information to be found in the details of acquisition, which has been virtually ignored in the comparative literature. For example, a great deal of information about how children manage series can be obtained by analysing the sequential characteristics of acquiring the series. Particular attention should be paid to the acquisition profile of individuals and their resistance points along the way to criterion.

The versatility of this paradigm for exploring these and other questions has yet to be fully explored. Our second test phase is a beginning. We have used the method of delays to explore ordinal position and phrasing effects. In this way we were able to get a hook into how the series is

organised and represented by young children without the problem of reducing the degrees of freedom as in "gap-skipping" test trials.

Set size should be considered as a variable. It has already been demonstrated that pigeons are constituent limited at four items or so. In monkeys, a five item set has been used as a standard. More interesting is the question of how many items can subjects handle. At what number of items will monkeys begin to show the effects of constituent limitation? In the human case, constituent limitation can be explored from a developmental point of view. At what ages do children show which levels of constituent limitation?

Whilst serial ordering is an interesting topic of study, its value for providing insights about the development of intelligence is likely to be limited. Serial order tasks may be regarded as accessing a skill which is low level. The fact that pigeons have been shown to manage information via serial order, indicates that it is likely to be a "design primitive" or basic structure rather than a higher order benchmark of cognitive achievement (McGonigle 1987). Arbitrary lists can be managed using unsophisticated means of control, in particular, search strategies dependent upon knowledge of a few crucial items (e.g. the initial and terminal items). Of greater interest for the study of cognitive development is discovering which types of series facilitate or resist higher means of control. Consider the contrast between strings which are made up of arbitrary items, such as colours, and strings which are made up of non-arbitrary items, such as boxes which monotonically increase in size. The arbitrary colour string is held together purely by the temporal cement between items. There is no other reason why one colour should follow another. In a monotonic size string however, there is an internal "semantics." The only way to control the colour string is by temporal phrasing, however in the size string there exists the possibility for dimensional control. Dimensional control brings with it the possibility of optimizing search through the set and has the advantage of increasing constituent capacity. Using non-arbitrary strings such as these, there is the potential to explore contrasts between control by temporal phrasing versus dimensional control and, most importantly, how dimensional control emerges in development. The ability of young subjects to control both arbitrary and non-arbitrary series is the focus of the next chapter.

Chapter 5: Training and resistance points in robustly failing subjects

Rationale

As a result of the previous experiment on the serial ordering abilities of four year olds (Experiment 2), we know that the touch screen apparatus and paradigms are a viable and fruitful means of investigating serial order with young subjects. Therefore we are licensed to recruit the touch screen apparatus and methods for further exploration. In the following experiments we use the touch screen to investigate the serial and seriation abilities of subjects younger than four years.

Whereas in the previous investigation subjects *older* than age four years were used, in the following experiments we concentrated on subjects *younger* than age four years. This particular age group was chosen for a number of reasons. In the extended clinical investigation it was found that spontaneous success or success with assistance was possible for some four year olds, whereas subjects younger than age four were found to be completely resistant to success at seriation tasks and efforts to ameliorate their failure. However, because these young subjects were found to have an inadequate task characterisation, using the new touch screen procedures it might be possible to better communicate the task to these young subjects. Young subjects' difficulty with state-based error analysis might be overcome by communicating error information directly through immediate feedback. Parsed input might further assist task understanding and allow for the exploration of set size as a variable. In this way, it may be possible to discover what set size capacity subjects have and where their resistance points are. In a sense, there was "everything to gain" in further exploring the dimensions of failure in subjects younger than four years since their performance was heretofore impenetrable. Therefore, we focussed on this younger group, concentrating on reasons underpinning their failure. In this way, we might arrive at a "baseline" of seriation ability in young children.

If young subjects achieve success at seriation through training, we may conclude that the primary sources of failure are the modeling and state-based error analysis requirements of the classical seriation task. If they do not succeed through training, we may conclude that failure in these subjects is well and truly robust. In this case, deeper, more basic sources of failure must be sought.

In Experiment 2, the focus was on the serial control of arbitrary lists (of colours). In the following experiments, the focus is shifted from the serial control of arbitrary lists to the control of non-arbitrary, serialable lists. This change was made for a number of reasons. Firstly, whereas tasks requiring the serial control of arbitrary lists are likely to be tapping into a process which is low level, tasks involving non-arbitrary lists may be capable of facilitating higher means of control. In this case, the particular type of non-arbitrary list used is a series of items related monotonically along the dimension of size (as per seriation tasks). Such lists afford the possibility of dimensional control, that is, the use of dimensional relationships inherent in the string to control and optimize search through the set. The benefit of dimensional control is that it allows subjects to control large sets. Because subjects need remember only their present position in the string and the relationship which holds between items, there is little strain on memory. With the arbitrary string however, the subject must rely on serial control alone, and is thus limited by memory constraints.

By comparing subject performance on these two types of lists, it is possible to investigate how young subjects control their search through a list of items. Again (as found in the animal literature), quantitative limitations can point to qualitative differences. If young subjects are limited in the number of dimensionally related items they can control, this may indicate that they are using serial control alone to search through the set. In this sense, they would be "unaware" of the dimensional relationship, and treat each item as if unrelated to the others. If further item limitations are found in subjects' performance on arbitrary sets, then the root of subjects' failure may be a more basic problem of limitations on serial control.

The following experiments have three main goals. Firstly, we investigate young subjects' abilities to control non-arbitrary, monotonic size series to see whether seriation is trainable (using a touch screen analog of the

seriation task) in three year old subjects who fail at a conventional version of the seriation task. Secondly, we aim to discover what kind of performance limitations exist, especially in terms of the number of items subjects can control.¹ Finally, we seek to discover the source of subjects' performance limitations with respect to serial control; are subjects' limitations due to problems with managing dimensional stimuli or a deeper problem with serial ordering *per se*?

EXPERIMENT 3.1

Subjects

7 children at the Edinburgh University Psychology Department Nursery, aged from 2;11 to 3;8 (years, months) served as subjects. This particular age group was chosen based on the age constraints of the "robust failure" category of performance found in the extended clinical investigation. Failure on a five-item (blocks) version of the seriation task was the criterion for selection in order to confirm subjects seriation failure status. A five-item set was chosen based on preliminary data that failure to seriate five items is consistent with the "robust failure" category of performance. A heavy investment was made in small number of subjects so that each subject could be treated in depth.

Apparatus

Subjects were individually tested in a small quiet room. Subjects were seated on a height-adjustable chair at a table in front of a computer screen with the experimenter seated beside the subject. The programming of trial events, stimulus presentation, and data recording were controlled by a BBC microcomputer equipped with a touch-sensitive screen measuring 20 cm by 27 cm. The stimuli were displayed in a horizontal line across the middle of the screen.

¹Note that by this we do not mean "partial seriation" with respect to a larger set (Piaget's "uncoordinated sets"). Here we ask how many items can subjects seriate without having a particular number as a "total" or "standard" (however due to technological constraints we were limited to a maximum of five items in this study).

Stimuli

Training stimuli consisted of five squares of the same colour with regularly differing sizes. The largest item measured 4 cm by 4 cm, the second biggest measured 3.5 cm by 3.5 cm, the middle item measured 3 cm by 3 cm, the second smallest measured 2.5 cm by 2.5 cm, and the smallest item measured 2 cm by 2 cm. The stimuli were colour coded for direction of search. Subjects searching from biggest to smallest (monotonic descending) trained with red stimuli. Subjects searching from smallest to biggest (monotonic ascending) trained with green stimuli.

Test stimuli in the over-training phase were identical to the training set except for the following differences in size. The second biggest item from the training set served as the largest item in the test set, the middle item from the training set was the second biggest item in the test set, and so on. The smallest item in the test set measured 1.5 cm by 1.5 cm.

Design

The training paradigm used is the "combined" paradigm. It is similar to the serial order paradigm used in Experiment 2, however it demands the combined use of both sequential and seriation components (see McGonigle, 1987). The subject must select (using the touch screen) a particular sequence of objects, in this case, squares which increase (or decrease) in size monotonically. Subjects were presented the stimuli in the "forward" training direction, that is, as they were to be reported in the final sequence (the first item first, the second item second, etc). Each subject was trained on a monotonic size series; either monotonic ascending or monotonic descending.

A within subjects design was used. Subjects were brought through the training and test procedures at their own pace. We began with the combined paradigm using monotonic size stimuli, rather than first assessing the full complement of individual components in isolation (as per the decomposition, see McGonigle, 1987; McGonigle & Chalmers, 1986; in press). This decision was taken as a shortcut in order that within subject analysis could be kept to manageable proportions. Using a failure-

based approach, performance limitations were then examined specifically to try and diagnose their source.

The experiment was divided into five main parts: pre-training, training, verbal probe, over-training, and delayed response testing. At each stage subjects were required to reach a performance criterion before going on to the next procedure.

SUBJECT	AGE	STRING
ALEXA	3,8	DSC
YVONNE	3,5	ASC
GILL	3,5	DSC
ANDREW	3,4	DSC
CATH	3,0	DSC
ZACH	3,0	ASC
NICOLE	2,11	ASC

Table 5.1: Subjects' age and string assignment

Procedures

Pre-Training. The subjects were introduced to the touch screen environment via the pre-training program used in Experiment 2 (Chapter 4). Subjects were questioned to ensure that they could distinguish the sizes of items displayed on the screen.

Training. Training proceeds as in Experiment 2. (a) First only one stimulus was presented and the subject was required to touch the stimulus for six successive trials. When the subject touched the stimulus, a registration tone sounded, followed by a feedback tone indicating a correct response (in the case of one item there are no degrees of freedom, and therefore, a correct response is the only response a subject can make). There was a five second interval between trials. A warning tone indicated that a new trial was about to begin. The position of the stimulus changed randomly after each trial. After subjects made six successive correct responses, they advanced to the next stage. At the bottom of the screen, the "analog display" allowed subjects to keep track of their performance. (b) A second stimulus was added (according to the sequence) and the

subject must report the items in the correct monotonic order. Whenever a subject touched a stimulus, a registration tone sounded. In the case of an incorrect response, the registration tone was immediately followed by a feedback tone, the screen went blank, and the trial was repeated after the usual five second intertrial interval. If four successive errors were recorded, the subject returned to the previous stage (one stimulus). Six successive correct responses were required to return the subject to two stimuli. Correct feedback tones were only given at the completion of a fully correct string. To prevent any cues from spatial position, each element changes location after each correct trial. Again a criterion of six successive correct responses were required for advancement to the next stage. (c) A third stimulus was added. The subject proceeded as above. If four successive errors were recorded, the subject returned to the previous stage (two stimuli). Six successive correct responses were required to return the subject to three stimuli. In turn, six successive correct responses were required for advancement to the next stage. (d) As before, with four stimuli. (e) Finally the subject was confronted with the entire set of five sizes. After subjects reached criterion with the five-item set they followed the verbal probe and over-training procedures (below).

Each subject was given one training session per day. Subjects always began a day's training with one item, restating each criterion regardless of where they had left off the previous day, until they reached the final five-item ABCDE criterion. Sessions were given on consecutive days as far as was possible ("school days" only). Sessions did not exceed 15 minutes in length and a session was terminated if a subject showed signs of tiring. A typical training session consisted of 30 trials (excluding correction trials).

Verbal Probe. Verbal probe questions were asked (in the same session) immediately after subjects reached the training criterion with the full five-item (ABCDE) set. Subjects were turned away from the screen so that it was no longer visible. They were then asked the following questions.

- A. (Prompted) What do you call the block you touch first?
(pause) and then (the next one)? (pause) and then? (pause)
and then? (pause) and then? (pause).
- B. (Prompted) What do you call the block you touch last?
(pause) and the one before that? (pause) and the one before

that? (pause) and the one before that? (pause) and the one before that? (pause).

Subject responses to these questions were recorded both on audio tape and protocol sheets ticked off in situ. No differential feedback was given, however subjects were given neutral praise and encouragement. Completion of the last question was followed by an invitation to "play a new game tomorrow." The session was terminated and the subject returned to the classroom.

Over-training. In order to check the reliability of performance, once subjects reached criterion with the entire set of five stimuli, in a separate session they were confronted with the entire set of five stimuli from the start. Randomly interspersed with ABCDE over-training trials were ABCDE trials with the test set of stimuli. The ratio of training set trials to test set trials was 50/50. Registration and feedback tones were the same as in the training phase. Criteria were as follows. First the subject was required to restate the training criterion of six correct trials in a row with (minimally) three trials from the training stimulus set and three trials from the test stimulus set. After meeting this criterion, the subject was required to meet a routinization criterion of nine correct trials out of ten with (minimally) four trials from each stimulus set. After reaching the routinization criterion, the subject moves on to delayed response testing.

Delayed Response Testing. Subjects were confronted with what appeared to be a normal five-item series (training stimulus set). Whilst the subject was responding however, the screen was cleared at quasi-random points in the series, i.e. sometimes after a response to the first item, sometimes after responses to the first two items, the first three, etc. The items were re-presented after a delay, during which a "bridging tone" was sounded. Upon re-presentation, items retained the same spatial locations as in the "priming" sequence. When the items were re-presented, the subject was required to "tell the machine" (by touching) which items had not been touched on the first presentation. An incorrect response resulted in the sounding of the appropriate feedback tone. The trial was terminated and re-presented as a correction trial (i.e. from the beginning of the trial, before the delay). The initial delay was two seconds. A criterion of four successive correct responses was required for the subject to increase the

delay by half (0.5) a second. Four successive errors resulted in a new trial in the two-second delay case (i.e. a new configuration of stimuli and a new delay position). In the case of delays longer than two seconds, the subjects delay was reduced by half a second.

The first four trials were "shaping" trials. In these trials, the display was not cleared. The experimenter interrupted the subject by holding back the subject's hand at a quasi-random position in the series. After two seconds, the subject was allowed to continue reporting the series. Subjects were given the following instructions for the remaining trials: "The machine is going to be tricky this time and may take away some of the squares, then the machine will bring them back again and you starting touching the ones the machine didn't let you touch before."

Results

Monotonic Size Series Acquisition

Item Limitation. Acquisition data are summarised in Table 5.2. Of our seven subjects only four were able to reach the five-item ABCDE criterion. Of the remaining three, Cath was limited to three items, Zach to four items, and Nicole to 2 items. Note that the subjects who reached the full ABCDE criterion were on average six months older than those subjects who were item limited to less than five.

			SET SIZE	TRIALS TO	TRIALS TO	TRIALS TO	TRIALS TO	
SUBJECT	AGE	STR	LIM.	AB	ABC	ABCD	ABCDE	TOTAL
ALEXA	3,8	dsc	5	16	15	13	12	56
YVONNE	3,5	asc	5	7	7	6	42	62
GILL	3,5	dsc	5	16	31	12	57	116
ANDREW	3,4	dsc	5	7	19	13	13	52
CATH	3,0	dsc	3	6	20	X	X	26
ZACH	3,0	asc	4	9	13	45	X	67
NICOLE	2,11	asc	2	12	X	X	X	12
TOTALS		dsc		45	85	38	82	250
		asc		28	20	51	42	141
		total		73	105	89	124	391
MEANS		dsc		11	21	13	27	63
		asc		9	10	26	42	47
		total		10	18	18	31	56

Table 5.2: Trials to criterion: size series training

Trials to the ABCDE criterion. Of the subjects who achieved the full five item criterion, there was a wide variation in number of sessions required to reach criterion. Alexa and Andrew required 5 and 4 sessions respectively, whereas Yvonne and Gill required 10 and 12 sessions respectively. In terms of trials to criterion Alexa, Yvonne, and Andrew required an average of 56 trials to reach the ABCDE criterion (excluding correction trials), whereas Gill required roughly double that number of trials (116).

Breakdown of criteria. Because the series was acquired in a parsed fashion, we can examine numbers of trials required to reach each criterion for each part of the series, i.e. the AB, ABC, ABCD, and ABCDE criteria. In this way we can locate resistance points in acquiring the series both for individuals and groups. Table 5.3 shows the percent of total trials to criterion required to reach each of the subseries criteria for each of the subjects who acquired the full ABCDE series. Profiles were for the most part an individual matter. Yvonne and Gill had the greatest difficulty in acquiring the fifth item (ABCDE). For Gill 68% of total trials to criterion were spent on ABCDE with the remaining trials distributed evenly across the other

criteria. For Yvonne 49% of total trials to criterion were spent on ABCDE. Another site of difficulty for her was ABC (27%). ABC was the site of greatest difficulty for Andrew, requiring 37% of trials. Alexa, unlike the other subjects, spent the largest portion of her trials (29%) on the AB criterion and then required steadily fewer trials for each successive criterion (27%, 23%, 21%).

			SET SIZE	TRIALS TO	TRIALS TO	TRIALS TO	TRIALS TO	
SUBJECT	AGE	STR	LIM.	AB	ABC	ABCD	ABCDE	TOTAL
ALEXA	3,8	dsc	5	29%	27%	23%	21%	56
YVONNE	3,5	asc	5	11%	11%	10%	68%	62
GILL	3,5	dsc	5	14%	27%	10%	49%	116
ANDREW	3,4	dsc	5	13%	37%	25%	25%	52
TOTALS				16%	25%	15%	43%	286
MEANS				16%	25%	15%	43%	72

Table 5.3: Distribution of trials to criterion: size series training

Performance of subjects who did not reach the ABCDE criterion. Subjects who did not reach criterion on the full ABCDE set showed a pattern of requiring more trials with the addition of each item. After acquiring the AB series, Nicole persevered for three sessions on the ABC series without any success. Cath required 3 times as many trials to acquire the ABC criterion than the AB criterion. She became frustrated with the four-item ABCD series after one session and refused to participate further.² Zach's effort in terms of trials to criterion was distributed 13% to acquire AB, 19% to acquire ABC, and 67% to acquire ABCD. He required 3.5 times more trials to acquire ABCD than ABC. He was exposed to the ABCDE series in one session after reaching the ABCD criterion, but without any success.

Error Type. Errors were divided into three types: forward errors, backward errors, and reference errors. Analysis of errors accumulated over sessions of training on ABCDE trials revealed that 85% of errors were of the

²This subject's lack of cooperation was likely to be part of a general disruption in her behaviour due to the arrival of a new sibling at home.

forward type, 11% were backward errors, and 4% were reference errors. Further breakdown of forward errors showed that 92.5% were of the one-step variety and 7.5% were of the two-step variety. One-step forward errors were also the most common type of error in ABC (65%) and ABCD (81%) training. However, there was also a sizeable proportion of reference errors (29%) in ABC training.

Error by Ordinal Position. In both ABCD and ABCDE training errors were evenly distributed across the middle items with very little error at the initial and terminal positions. In ABCD training 8% of error occurred at the first position (A). All (and only) reference type errors occur at the first position. 42% of errors occurred at the second position (B), 46% at the third position (C), and 4% at the final position (D). According to the training procedures used in this experiment only a backward error can occur at the final position. In ABCDE training 4% of error occurred at the first position (A), 30% of errors occurred at the second position (B), 34% at the third position (C), 30% at the fourth position (D), and 2% at the final position (E).

Verbal Probes

The four subjects who reached the ABCDE criterion were asked questions about the series. Results can be found in Table 5.4. Subjects verbal command of the series was not good, however all subjects except Gill made some attempt to use size terms to talk about the series. Only Alexa was able to name the sizes using a variety of gradable terms: big, middle, little, and tiny. However, she ran out of terms and returned to "big" again. Yvonne, on the other hand, had two categories for the items. She classified the first three items as "little" and the last two as "big." Andrew seemed to be trying to say as much as possible about the items (red, wee, big), but not in any principled way. No subject could correctly name the size when prompted from E. Alexa only reported the last one as "the little one." Yvonne referred to the last item as "enormous" and to the next one as "big," but then ran out of terms.

SUBJECT	ALEXA dsc	YVONNE asc	GILL dsc	ANDREW dsc
ABCDE A	the big one	little	don't know	the red one
ABCDE B	middle one	little		the wee one
ABCDE C	the little one	little		the big one
ABCDE D	the tiny one	big		
ABCDE E	the big one	big		
EDCBA E	the little one	enormous	don't know	don't know
EDCBA D	"no more"	big		
EDCBA C		enormous		
EDCBA B		"no more"		
EDCBA A				

Table 5.4: Size series verbal probes

Over-training and Stimulus Size Set Test

Performance. All subjects who acquired the ABCDE criterion (Alexa, Yvonne, Gill, and Andrew) were given at least one session of over-training trials interspersed with test trials. Each subject was given three sessions to achieve the criteria before they were removed from the task. All subjects except Gill were able to restate the six correct trials in a row acquisition criterion. Only Alexa and Andrew were able to reach the nine correct trials out of ten routinization criterion. Overall, subjects were 67% correct on training stimulus set trials and 69% correct on test stimulus set trials, however there was considerable individual variation. Both Alexa and Yvonne performed better on training trials (18 and 10 percentage points respectively). Alexa was 91% correct on training trials and 73% correct on test trials. Yvonne was 74% correct on training trials and 64% correct on test trials. Opposite to Yvonne, Andrew's performance on test trials was 10 percentage points better than his performance on training trials (77% and 67% correct respectively). Gill also showed better performance on test trials (58% correct) than training trials (33% correct), however her performance was poor in general.

Error Type. Analysis of errors accumulated in over-training revealed that for training stimulus set trials 52% of errors were of the forward type, 30% were backward errors, and 17% were reference errors. For test stimulus set trials the distribution of error types was roughly the same: 50% of errors

were of the forward type, 32% were backward errors, and 18% were reference errors. There was a large increase in the proportion of backward and reference errors compared to the proportion of backward and reference errors in training; a 54 percentage point *decrease* in forward errors, a 20 percentage point increase in backward errors, and a 14 percentage point increase in reference errors.

Error by Ordinal Position. In training set trials, errors were distributed 17% at the first position (A), 9% of errors occurred at the second position (B), 39% at the third position (C), 26% at the fourth position (D), and 9% at the final position (E). In test set trials, 18% of error occurred at the first position (A), 23% of errors occurred at the second position (B), 32% at the third position (C), 14% at the fourth position (D), and 14% at the final position (E). These patterns of distribution of error by ordinal position were different from that which was found in training, where error was negligible at the poles and evenly distributed over the three middle items. Similar to training however, the greatest amount of error occurred at C, the middle item. The greater amount of error at A was due to the increase in reference errors. In 75% of these errors (for both the training and test stimulus set) the subjects chose B rather than A.

Decision Time. Inter-item response time data were taken from the correct trials which made up the two criteria. In cases where this data was not available, the data were taken from correct trials within the over-training phase. The mean response latency to each item of the series was calculated for each subject. Values greater than or equal to 3 times the standard deviation were replaced by the mean. Average decision times for completion of the entire ABCDE series can be found in the end column of Table 5.5. The average decision time was 8.49 seconds, with 8.51 seconds for the training set trials and 8.48 seconds for the test set trials. All subjects except Andrew spent (on average) between one half to three quarters of a second less time to complete test set trials than training set trials. Andrew, on the other hand spent (on average) a second and a half more time to complete training set trials than test set trials.

Training Stimulus Set

Subject	A	B	C	D	E	Total
Alexa	309 (31)	221 (22)	191 (19)	153 (15)	127 (13)	1002
Yvonne	203 (27)	174 (23)	134 (18)	143 (19)	104 (14)	757
Gill	169 (17)	115 (12)	215 (22)	292 (30)	185 (19)	975
Andrew	150 (22)	149 (22)	135 (20)	126 (19)	111 (17)	670
MEAN	208 (24)	165 (19)	169 (20)	178 (21)	132 (15)	851

Test Stimulus Set

Subject	A	B	C	D	E	Total
Alexa	261 (27)	206 (21)	181 (19)	189 (20)	126 (13)	964
Yvonne	193 (28)	155 (22)	130 (19)	122 (18)	96 (14)	696
Gill	205 (23)	158 (18)	208 (23)	164 (18)	168 (19)	902
Andrew	177 (21)	210 (25)	152 (18)	127 (15)	162 (20)	828
MEAN	209 (25)	182 (22)	168 (20)	151 (19)	138 (16)	848

Table 5.5: Mean decision time and IRT expressed in hundredths of seconds (% of decision time in parenthesis)

Inter-item RT. Table 5.5 shows the average inter-item RT for all subjects. For training stimulus set trials, mean inter-item RT was 2.08, 1.65, 1.69, 1.78, and 1.32 seconds respectively and for test stimulus set trials 2.09, 1.82, 1.68, 1.51, and 1.38 seconds respectively. The pattern of latencies for individuals, except Gill, generally followed a pattern a steady decline in time needed to report each item for training set trials. That Gill had a different inter-item RT profile from other subjects may be linked to her poor performance *vis a vis* other subjects. For all subjects except Gill, the greatest percentage of total decision time in training set trials was spent responding to A (27%), declining a few percentage points with each additional item (22%, 19%, 17%) until reaching E (14%) (see Figure 5.1). If the "expense" of inter-item RT is expressed in terms of ranks (1 = least time, 5 = most time), we find that for both the training set and test set trials A = 5, B = 4, C = 3, D = 2, and E = 1. Analysis using Kendall's coefficient of concordance showed that this ranking ($W = 0.96$) was significant at the $p < 0.02$ level. The pattern of inter-item RT for test set trials showed a great deal of individual variation. However, subjects did not show any clear evidence of phrasing.

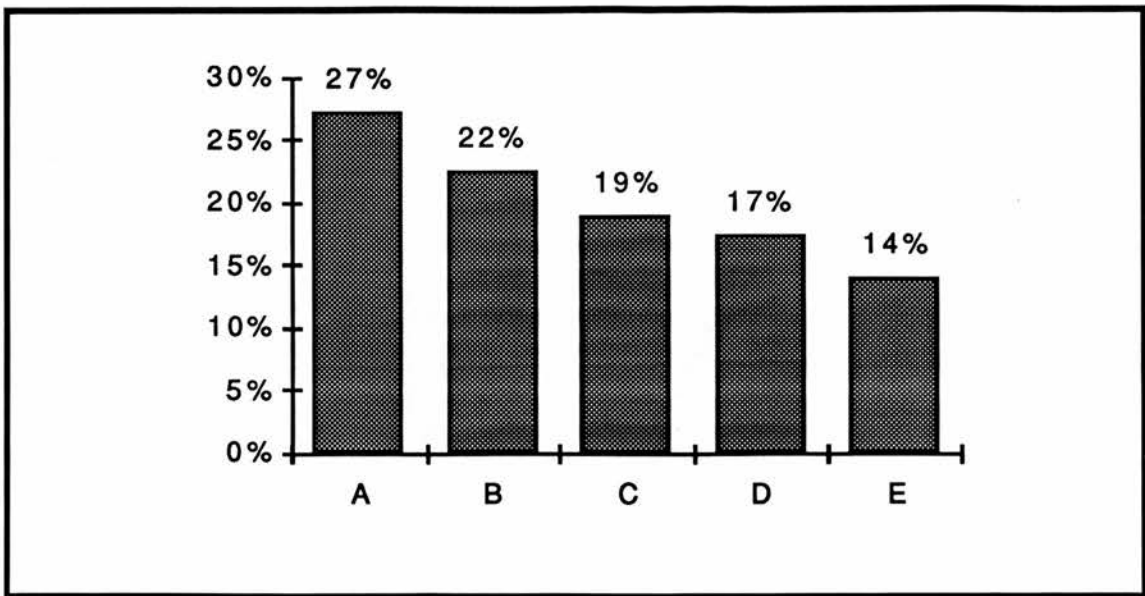


Figure 5.1: Mean IRT on training set trials as a % of decision time (all subjects except Gill)

Delayed Response Testing. Only subjects who had met the “nine correct trials out of ten” criterion in the over-training phase (Andrew and Alexa) participated in this task. Although these subjects were able to correctly complete all of the “shaping” trials, they performed very poorly on the task proper. After one session each, neither subject had logged a single correct trial. Four out of six of Andrew’s errors were to begin the series again from the first item, regardless of the position of the delay. Alexa’s errors showed no particular pattern. Both subjects were returned to the over-training task until they restated the “nine correct trials out of ten” criterion. Alexa was unable to restate this criterion. Andrew accomplished this in three sessions and was returned to the delay task beginning with a reduced, half-second delay. Again there was complete failure. The subject always returned to the first item after the delay regardless of the position of the delay.

Results Summary

Our three year old subjects had great difficulty learning a five-item size series. Whilst four subjects were able to reach criterion with the five-item set, only two were able to maintain stable performance. Failure to complete any particular size series was primarily due to the child’s tendency to skip one-step forward in the sequence. Errors were most likely to occur in the middle of a series rather than at the initial or terminal

positions. Subjects possessed very limited ability to express knowledge about size series verbally. If they could say anything at all about the sizes they had been reporting, they generally classified them as either "big" or "little."

In over-training trials, only two subjects were able to reach the nine correct trials out of ten over-training criterion. Most achieved a higher level of performance with the training set of stimuli than with a test set of stimuli. Subjects took about eight and a half seconds to complete a trial, regardless of stimulus set. Analysis of inter-item RT showed that the greatest amount of time was spent responding to A and there was steady decline in time needed to report an item thereafter. In delayed response testing, subjects were unable to complete a series despite a very short (half second) delay.

Discussion of Experiment 3.1

Seriation failure of three year old subjects was confirmed. With protracted training, despite rapid, if limited, learning of partial size series, most subjects were unable to reach the five-item criterion. For most subjects stable performance was possible only with small sets of items, with four items being an upper bound. Thus we see the seriation failure of pre-four year olds confirmed, despite a training-intensive, non-verbal paradigm. Furthermore, subjects who did manage to reach the five-item criterion were easily "put off" successful performance.³ The change in the absolute sizes of items in the over-training phase caused deficits in performance of up to 18%. Also, the number of reference errors at positions A and B increased, reflecting confusion at the point where the two stimulus sets overlapped in size. Any interruption in the series, even as short as half a

³The instability of seriation performance was also informally confirmed in a series of preliminary block seriation retests which followed touch screen training. Although most subjects were now able to achieve trial and error success at seriation with five items, this "success" was neither consistent nor reliable. Subjects could easily be "put off" success by spatial factors such as whether the model had the biggest item on the left of the right or by which block the subject chose first. Even the colour of the blocks might make a difference. "Successes," when they occurred were likely to be the result of practice and familiarity with blocks. Tellingly, no subject was able to achieve even fleeting success at seriation with ten blocks, nor did subjects use principled means to construct block series.

second, produced a complete breakdown in performance. This may be interpreted as indicating that subjects needed the first item as a "starting block" for completing the series. Intolerance of delay was also observed informally in training sessions. If subjects were distracted or paused to comment at any point after they began reporting the series, they were typically unable to continue without error. When they returned to reporting the series, they most typically began again with the first item. Subjects' abilities to verbally describe the gradient nature of a monotonic size series were also very limited. Their use of only two terms to describe the items ("big" and "little") may indicate that they have binned the items into two categories ("big ones" and "small ones") rather than as gradable from biggest to smallest.

What is the source of subjects' performance limitations? Is the cause of limitations in performance an inability to order items along a particular dimension (in this case: size), or is it the more basic problem of a limitation of serial control? In order to answer this question, in the next experiment we assess our subjects' ability to serially order unrelated stimuli (coloured squares of the same size) using the serial order paradigm (see Experiment 2). If subjects do not show similar performance limitations with unrelated stimuli, then we may conclude that it is not serial ordering ability *per se* which is causing the default. Rather, it is an inability of subjects to use the size dimension to control and assist their search through a series. If subjects show similar performance limitations, then we may conclude that performance limitations are *at least* limitations of serial control.

EXPERIMENT 3.2

Subjects

The same children used in Experiment 3.1 served as subjects. The only exception was Cath, who was dropped from the study at this time due to a lack of cooperation.

Apparatus & Stimuli

The apparatus was the same as used in Experiment 3.1. The stimuli consisted of five 3.5 cm by 3.5 cm squares of different colours (red, green,

yellow, blue, pink). The stimuli were displayed in a horizontal line across the middle of the screen, separated by 1.5 cm.

Design & Procedures

Training. Subjects were assigned to one of two colour sequences. To ease introduction to the colour series, subjects' first colour in the series was the same as the colour of stimuli they had used in Experiment 3.1. For subjects in the monotonic ascending group the colour sequence was green, red, yellow, blue, pink. For subjects in the monotonic descending group the colour sequence was red, green, yellow, blue, pink. All subjects were trained in the forward training condition, i.e., subjects were presented the stimuli as they were to be reported in the final sequence, the first item first, the second item second. etc. Training proceeded as in Experiment 3.1.

Verbal Probe. Verbal probe questions were asked (in the same session) immediately after subjects reached the training criterion with the full five-item (ABCDE) set. Subjects were turned away from the screen so that it was no longer visible. They were then asked the following questions.

A. (Free Recall) Can you tell me which colours you touched on the screen? (The experimenter pauses to allow the subject to freely recall the colours touched on the screen).

B. (Prompted) Which colour do you touch first? (pause) and then? (pause) and then? (pause) and then? (pause) and then? (pause).

C. (Prompted) Which colour do you touch last? (pause) and before that? (pause) and before that? (pause) and before that? (pause) and before that? (pause).

D. (Gap Filling) Imagine/Suppose/If you had only two colours on the screen, and one was D (the experimenter fills in the appropriate colour for the subject) and the other was B. Which one would you touch first? (pause) Which one would you touch last? (pause)

E. (Exclusion) If B and D are on the screen, which colours are left out/not there/missing?

Subject responses to these questions were recorded both on audio tape and protocol sheets ticked off in situ. No differential feedback was given, however subjects were given neutral praise and encouragement. Completion of the last question was followed by an invitation to "play a new game tomorrow." The session was terminated and the subject returned to the classroom.

Over-training. In order to check the reliability of performance, subjects reached criterion with the entire set of five stimuli, in a separate session they were confronted with the entire set of five stimuli from the start. Subjects were required to reach a criterion of nine correct trials out of ten.

Results

Arbitrary Colour Series Acquisition

Item Limitation. Acquisition data are summarised in Table 5.6. Of our six subjects only three were able to reach the five-item ABCDE criterion. Of the remaining three, Gill was limited to four items, Zach to three items, and Nicole to two items.

		SET SIZE	TRIALS TO	TRIALS TO	TRIALS TO	TRIALS TO	
SUBJECT	AGE	LIM.	AB	ABC	ABCD	ABCDE	TOTAL
ALEXA	3,8	5	6	6	17	11	40
YVONNE	3,5	5	6	6	6	11	29
GILL	3,5	4	15	6	24	X	45
ANDREW	3,4	5	6	15	6	11	38
ZACH	3,0	3	9	6	X	X	15
NICOLE	3,0	2	11	X	X	X	11
TOTALS			53	39	53	33	178
MEANS			9	8	13	11	30

Table 5.6: Trials to criterion: colour series training

Trials to the ABCDE Criterion. Of the subjects who achieved the full five item criterion, there was a little variation in number of sessions required to reach criterion. Yvonne and Andrew required two sessions, whereas Alexa required only one session. In terms of trials to criterion Alexa, Yvonne, and Andrew required an average of 36 trials to reach the ABCDE criterion; 40, 29, and 38 trials respectively.

Breakdown of criteria. Table 5.7 shows the percent of total trials to criterion required to reach each of the subseries criteria for each of the subjects who acquired the full ABCDE series. Profiles were for the most part an individual matter. Alexa spent the largest portion of her trials (43%) on the ABCD criterion. AB and ABC were acquired with the minimum number of trials to criterion. Yvonne had the greatest difficulty in acquiring the fifth item (ABCDE). 38% of total trials to criterion were spent on ABCDE with the remaining trials distributed evenly across the other criteria (21% each). AB, ABC, and ABCD were acquired with the minimum number of trials to criterion. ABC was the site of greatest difficulty for Andrew, requiring 39% of trials. Another site of difficulty was ABCDE, requiring 29% of trials. AB and ABCD were acquired with the minimum number of trials to criterion.

		SET SIZE	TRIALS TO	TRIALS TO	TRIALS TO	TRIALS TO	
SUBJECT	AGE	LIM.	AB	ABC	ABCD	ABCDE	TOTAL
ALEXA	3,8	5	15%	15%	43%	28%	40
YVONNE	3,5	5	21%	21%	21%	38%	29
ANDREW	3,4	5	16%	39%	16%	29%	38
MEANS			17%	25%	27%	31%	36

Table 5.7: Distribution of trials to criterion: colour series training

Performance of subjects who did not reach the ABCDE criterion. After acquiring the AB series, Nicole worked on the ABC series for a further session without success. Zach acquired AB and ABC very efficiently (i.e., near the minimum number of trials to criterion), but was unwilling to continue training on the ABCD series due to the considerable effort

required for him to reach criterion on ABCD for the size series. Gill spent 4 sessions (24 trials) on the ABCDE series before she became unwilling to participate further.

Comparison with training results in Experiment 2 (Chapter 4). Table 5.8 summarizes acquisition data for subjects from the present study and older subjects (from the forward training condition only) in Experiment 2. All subjects three years eight months and older were able to acquire the full five-item ABCDE set. At age three years five months, subjects began to show limitations in the number of items they could control in a serial production.

			TRIALS TO	TRIALS TO	TRIALS TO	TRIALS TO	
SUBJECT	AGE	LIMIT	AB	ABC	ABCD	ABCDE	TOTAL
BARRY	4,7	5	N/A	N/A	N/A	N/A	N/A
JANE	4,6	5	6	6	7	11	30
BEN	4,3	5	5	5	9	6	25
LOUISE	4,1	5	6	6	12	11	35
LEWIS	3,10	5	6	6	19	8	39
KIRSTY	3,9	5	6	10	9	8	33
<i>Mean</i>			6	7	11	9	32
ALEXA	3,8	5	6	6	17	11	40
YVONNE	3,5	5	6	6	6	11	29
GILL	3,5	4	15	6	24	X	45
ANDREW	3,4	5	6	15	6	11	38
ZACH	3,0	3	9	6	X	X	15
NICOLE	2,11	2	11	X	X	X	11
<i>Mean</i>			9	8	13	11	30
MEAN			7	7	12	10	31

Table 5.8: Comparison with subjects in Experiment 2: trials to criterion

Although fewer subjects in the present study reached the ABCD and ABCDE criteria, mean numbers of trials to each criterion were roughly the same for both studies. However, it must be remembered that whilst subjects in Experiment 2 were “touch screen naive,” subjects in the present

study had already had experience of the touch screen and monotonic size series training. This familiarity with the apparatus and the paradigm may have reduced the number of trials they might have otherwise required to reach subseries criteria.

Table 5.9 shows the percent of total trials to criterion required to reach each of the subseries criterion for all subjects. Subjects in Experiment 2 required the largest percentage of trials to reach the ABCD criterion. The majority of these subjects required a minimum or near minimum number of trials to acquire both AB and ABC (a mean of 17% each). About 40% of trials are used to acquire ABCD and 26% to acquired ABCDE. Among subjects in the present study only Alexa (3,8) follows this pattern. However, Jane (4,7) and Yvonne (3,5) follow the same pattern: minimal trials to acquire AB, ABC, and ABCD, with most trials required to acquire ABCDE. Also both Kirsty (3,9) and Andrew (3,4) have particular difficulty with ABC. For the entire group, the difficulty of acquiring subseries criteria may be expressed in terms of ranks (1=least trials, 4=most trials): AB=1, ABC=2, ABCD=4, ABCDE=3. Analysis using Kendall's coefficient of concordance ($W = 0.50$) showed that this level of concordance in rankings was significant at the $p < 0.01$ level.

		TRIALS TO	TRIALS TO	TRIALS TO	TRIALS TO	
SUBJECT	AGE	AB	ABC	ABCD	ABCDE	TOTAL
JANE	4,6	20%	20%	23%	37%	30
BEN	4,3	20%	20%	36%	24%	25
LOUISE	4,1	17%	17%	34%	31%	35
LEWIS	3,10	15%	15%	49%	21%	39
KIRSTY	3,9	18%	30%	27%	24%	33
<i>Mean</i>		<i>18%</i>	<i>20%</i>	<i>35%</i>	<i>27%</i>	<i>32</i>
ALEXA	3,8	15%	15%	43%	28%	40
YVONNE	3,5	21%	21%	21%	38%	29
ANDREW	3,4	16%	39%	16%	29%	38
<i>Mean</i>		<i>17%</i>	<i>25%</i>	<i>27%</i>	<i>31%</i>	<i>36</i>
MEAN		18 %	22 %	31 %	29 %	34

Table 5.9: Comparison with subjects in Experiment 2: distribution of trials to criterion

Error Type. Analysis of errors accumulated over sessions of training on ABCDE trials revealed that 67% of errors were forward errors and 33% were backward errors. All forward errors were of the one-step variety. There was only one reference error in ABCDE training which was made by Gill, who spent 24 trials training on ABCDE but did not reach criterion. She also had the same proportion of error types as did successful subjects, however she made more errors more frequently. In ABCD training forward errors were the most common type of error at 80%. 60% were of the one-step and 20% were of the two-step variety. 20% of errors were reference errors. There were no backward errors. In ABC training errors were evenly divided between forward and reference errors.

Comparison with data from Experiment 2 revealed that the older group of subjects produced only forward errors in ABCDE training. The subjects in the present study's errors were 67% forward errors and 33% backward errors. However examination of individuals' error data revealed that Alexa's error profile conformed to that of the subjects in Experiment 2.

Error by Ordinal Position. In ABCDE training, errors were concentrated at the end of the series. 14% occurred at the third position (C), 71% occurred at the fourth position (D), and 6% at the final position (E). In ABCD training, errors were concentrated in the middle of the series. 20% of error occurred at the first position (A), 40% of errors occurred at the second position (B) and 40% at the third position (C). There were no errors in the final position.

Verbal Probes

The three subjects who reached the ABCDE criterion were asked questions about the series. Results can be found in Table 5.10. Similar to results found with older children, subjects' verbal command of the series was not good, however, all subjects except Andrew were able to freely name all the colours they had touched in the correct order. Andrew was unable to answer any of the verbal probe questions correctly. Only Yvonne was able to report all the colours correctly when prompted for them. Alexa reported ABCE, one of the most common training errors: a one-step forward error at C. No subject could correctly report the series when prompted from E. Although Yvonne did name all the colours

exhaustively, she did not report them in the correct order. Neither could subjects correctly answer the “gap-filling” and exclusion questions.

SUBJECT	AGE	COND	FREE (ABCDE)	PROMPT (ABCDE)	PROMPT (EDCBA)	GAP (BD)	EXCL. (ACE)
ALEXA	3,8	dsc	ABCDE	ABCE	E	BE	AE
YVONNE	3,5	asc	ABCDE	ABCDE	EDCAB	AE	A
ANDREW	3,5	dsc	DEAD	ABAE	don't know	DA	C

Table 5.10: Colour series verbal probes.

Over-training

Performance. All subjects who acquired the ABCDE criterion (Alexa, Yvonne, and Andrew) were given at least one session of over-training trials. Andrew and Yvonne were able to meet the nine correct trials out of ten criterion in one and two sessions respectively. Alexa was unable to meet this criterion after four sessions of over-training trials. Andrew and Yvonne were 83% and 79% percent correct respectively over all over-training trials. Alexa achieved only 50% correct.

Error Type. Analysis of errors accumulated in over-training revealed that 63% of errors were of the forward type, 17% were backward errors, and 21% were reference errors. There was a large increase of reference errors, whereas in training they were virtually nonexistent. Also there was an increase in two- and three-step forward errors which were not found in training. However, the most common type of error was (again) one-step forward errors which accounted for 42% of all over-training errors.

Error by Ordinal Position. Errors were more evenly distributed across ordinal position in over-training, with a peak in error at the middle position C. Errors were distributed 21% at the first position (A). 25% of errors occurred at the second position (B), 29% at the third position (C), 25% at the fourth position (D). There were no errors at the final position (E).

Decision Time. Inter-item response time data were taken from the correct trials which made up the nine correct trials out of ten criterion. In Alexa’s

case, where this data was not available, the data were taken from nine correct trials within the over-training phase. Average decision times for completion of the entire ABCDE series for can be found in the end column of Table 5.11. The average decision time was 13.06 seconds. There was a great deal of variation in decision time among individual subjects. Similar to results found with older children, Andrew and Alexa took roughly 11 seconds to complete the series. Yvonne, however, averaged 17 seconds. Most of the extra decision time required by Yvonne was spent responding to the first item. She required on average three seconds more to respond to the first item than did Andrew or Alexa.

Subject	A	B	C	D	E	Total
Alexa	253 (24)	244 (23)	244 (23)	190 (18)	142 (13)	1073
Yvonne	621 (36)	288 (17)	313 (18)	295 (17)	195 (11)	1712
Andrew	310 (27)	297 (26)	172 (15)	238 (21)	117 (10)	1134
MEAN	395 (30)	276 (21)	243 (19)	241 (18)	151 (12)	1306

Table 5.11: Mean decision time and IRT expressed in hundredths of seconds (% of decision time in parenthesis)

Inter-item RT. Table 5.11 shows the average inter-item RT for all subjects. The mean inter-item RT was 3.95, 2.76, 2.43, 2.41, and 1.51 seconds respectively. Individually, subjects spent the greatest proportion of decision time on the first item (mean 30%) and the least on the last item (mean 12%).

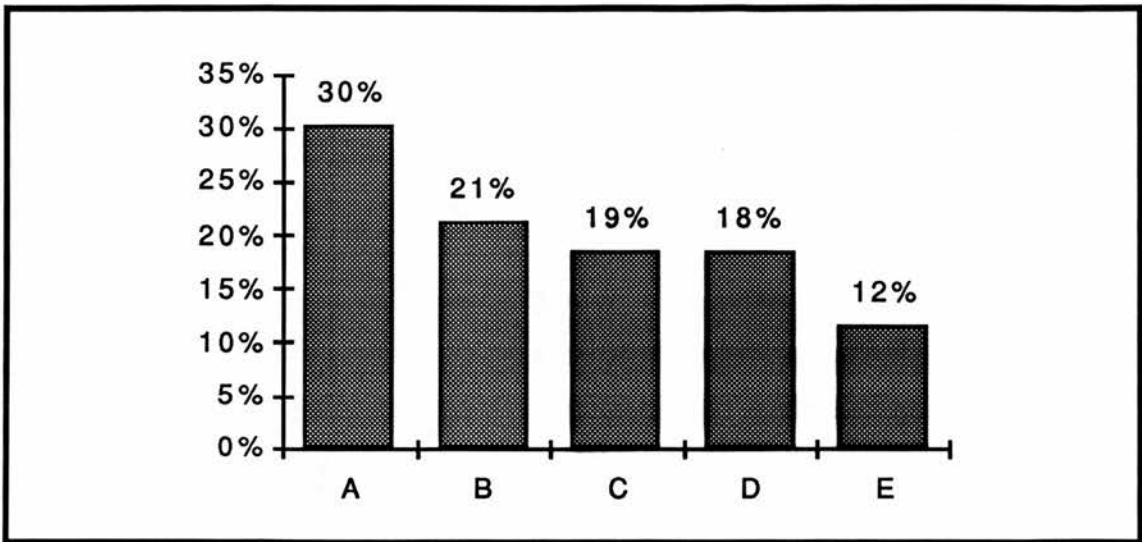


Figure 5.2: Mean IRT as a % of decision time (group).

However, the distribution of decision time over the middle items varied with the individual. Alexa spent roughly the same time for each of the first three items (24%, 23%, and 23%), but steadily less time was used to respond to items D and E (18% and 13% respectively). Yvonne required a great deal of time to respond to the first item (36%), spent roughly the same proportion of decision time on the next three items, and then showed a drop for the last item (11%). Andrew showed more “ups and downs” in his inter-item RT profile than other subjects. This pattern suggests that he may have been spontaneously reporting the ABCDE series in three phrases, A-BC-DE. The reader will recall that spontaneous phrasing was found in older children. In the forward training case the pattern was AB-CDE and in the backward training case the pattern was ABC-DE.

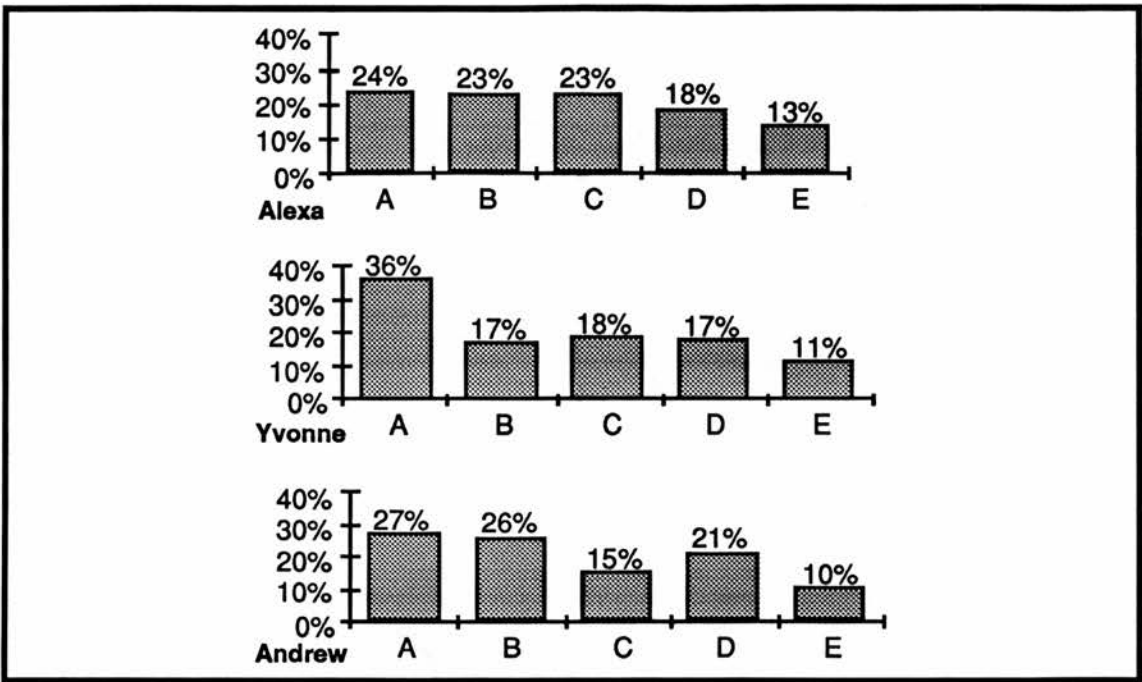


Figure 5.3: Inter-item RT as a % of decision time (individual subjects).

Comparison of Size Series and Colour Series Results.

Item Limitation. Acquisition data for both size and colour series are summarized in Table 5.12. Subjects showed similar performance for both series in terms of item limitation. In terms of both trials and sessions to criterion, subjects required fewer trials in colour series training to reach the same levels of performance as in size series training.

SUBJECT	SERIES	LIMIT.	SESSIONS	TRIALS	ERRORS
ALEXA	size1	5 items	5	56	20
	colour1	5 items	1	40	5
YVONNE	size1	5 items	10	62	20
	colour1	5 items	2	29	4
GILL	size1	5 items	12	116	44
	colour 1	4 items	7	45	8
ANDREW	size1	5 items	4	52	13
	colour 1	5 items	2	38	8
CATH	size1	3 items	2	26	5
	colour 1	SUBJECT UNCOOPERATIVE			
ZACH	size1	4 items	10	67	20
	colour 1	3 items	3	15	1
NICOLE	size1	2 items	5	12	2
	colour 1	2 items	3	11	2
TOTALS	size1		48	391	124
	colour 1		18	178	28
MEANS	size1		7	56	18
	colour 1		3	30	5

Table 5.12: Comparison of size and colour series

Breakdown of criteria. Most subjects showed similar patterns of distribution of trials to criterion across both tasks. These varied according to the individual. Yvonne, for example, in both size and colour series training required small percentages of trials to acquire AB, ABC, and ABCD, whereas the greatest investment in terms of trials is in learning ABCDE. Andrew spent the majority of total trials to criterion learning the ABC series in both size and colour series training. Alexa, on the other hand, showed different patterns of distribution of trials to criterion. In size series training she showed a steady decline in trials to criterion required for each additional item, whereas in colour series training she required a small percentage of trials for AB and ABC (15%), the most for ABCD (43%), and about halfway between these two for ABCDE (28%). This pattern was characteristic of that found for older subjects in Experiment 2.

Error. In ABCDE training for both series, one-step forward errors were the most common type of error. However, there were more backward errors in colour training than in size training (11% for size versus 33% for colour). There was also a difference in the ordinal position of error. In size series training, there was an effect of serial position, with error occurring most frequently in the middle position (C) with equally less

error occurring as a function of distance from the center. In colour series training however, almost all error occurred at C and D (14% and 71%) with none at A or B and very little at E (6%).

Decision Time and inter-item RT. Decision time for the completion of the ABCDE series in over-training were shorter for the size series than for the colour series. On average it took subjects 8.5 seconds to complete a size series and 13 seconds to complete a colour series. Individually the gap between time required to complete size and colour series varied. Whilst Alexa took only about one second longer to complete a colour series, Andrew took about four seconds longer, and Yvonne took about ten seconds longer. Whilst there was a general pattern of distribution of inter-item RT in size over-training trials, the pattern for colour over-training was dependent on the individual.

Results Summary

Most of our subjects were limited in serial ordering ability. Despite considerable experience of the apparatus and intensive training provided, only three subjects were able to reach criterion with the five-item set and only two were able to maintain stable performance. For most subjects stable performance was possible only with small numbers of items. Subjects showed the same "set size" restrictions whether stimuli were ordinally related along a size dimension (sized squares) or unrelated (coloured squares). Limitations observed for the colour series were, in almost every case, precisely the same limitations observed for size series. Comparison with older subjects in Experiment 2, showed that subjects show limitations in the number of items they can control in a serial production at age three years five months.

Failure to complete a colour series was primarily due to the child's tendency to skip one item forward in the sequence. Errors were most likely to occur in the middle of a series (C and especially D) rather than at the initial or terminal positions.

Subjects possessed very limited ability to express knowledge about colour series verbally. Two out of three subjects were able to name all the colours in the series, but only under the conditions of free classification. No subject was able to correctly report the series, in any direction, when

prompted for items. Neither were subjects able to correctly answer "gap-skipping" or exclusion questions.

In over-training trials, only two subjects were able to reach the nine correct trials out of ten over-training criterion. Subjects took about 13 seconds to complete a trial. Analysis of inter-item RT for Alexa and Yvonne did not reveal a phrasing effect. However the pattern of inter-item RT for Andrew suggests three phrases: A-BC-DE.

Discussion of Experiment 3.2

Three year old subjects were limited in their ability to report a series of unrelated elements. Despite considerable experience of the apparatus and intensive training provided, most of our three year old subjects were unable to report a series of five unrelated elements. Subjects were often limited to four items or less. Even for those subjects who could reach the criterion with five items, this was no guarantee that stable performance could be maintained. Furthermore, set size limitations were in nearly every case the same for both the ordinally related (monotonic size) strings in Experiment 3.1 and unrelated colour stimuli in Experiment 3.2. Subjects showed very similar performance on both tasks in terms of the number of items they were able to control. Again, the majority of subjects were limited to the control of four or fewer items. The similarity of these findings may be interpreted as indicating that subjects' failure at seriation is *at least* a failure of serial control. Although further assays are required in order to assess the full extent of why young children fail, having discovered that serial order is defaulting is an important achievement. In this respect, it is possible to say that one of the primary reasons that young children fail to seriate in the classical test, is that they are limited at controlling strings of any kind.

That five-items is enough to produce failure in many subjects is significant with respect to studies of serial learning with non-humans. In particular, this level of item limitation may be compared to that which is found with pigeons. In the work of Terrace and his colleagues we see that four- and five-item lists are taxing the limit of pigeon's abilities. The present data indicate that this may be the case for three year olds as well. As has been found with pigeons (see D'Amato & Colombo, 1988), that three year olds are able to reach any level of competence in serial order

with small sets may be due to "exclusion" strategies rather than an effective internal representation of the ABCDE series. With small sets, knowledge of item position, especially with respect to initial and terminal items, coupled with prudent "gambling" on internal items is sufficient to achieve high levels of performance. Beyond four or five items however, this kind of strategy comes up against too many degrees of freedom to support high levels of performance.

An example. Gill's errors in colour series training (Experiment 3.2) increase from one on ABC to five on ABCD to eight on ABCDE, where she did not reach criterion. In training on ABC, the vast majority of her errors occurred at B, AC errors being the most typical. Her response accuracy at A and C was 100%, however response accuracy at B dropped to 83%. We interpret this to mean that she had learnt that A was first and then she had only to resolve the conflict between B and C. C, being the terminal item, has special status as well. Therefore B would be a prudent choice after A. If any doubt was left about C, after a correct response to B then there was only C left to do. In this way, the subject could reach a high degree of accuracy by knowing the positions of item A, and perhaps C, only. In training on ABCD, Gill's response accuracies at A and D were 100%, however response accuracy at B dropped to 92% and further to 86% at C. Here there was conflict not only between B and C, but between C and D as well. The subject may have, through her experience of ABC, gained knowledge not only of the position of A, but could be fairly sure of B as well. The only conflict remaining was between C and D. Indeed the majority of errors were ABD. However, if the subject knew that D was the terminal item, she would chose C first. Then only D was left. However, now the subject needed to know the positions of A and B minimally and perhaps D as well. In training on ABCDE the subject did not reach criterion. By looking at the ordinal positions of her errors, we can see what may have gone wrong with her strategy. Gill's response accuracy at A was 96%, at B 100%, and at E 100%, however response accuracy at C dropped to 78% and to 89% at D. Here there were three conflicts, B v. C, C v. D, and D v. E. The subject probably had clear knowledge of positions of items A and B, but the two remaining conflicts between C and D, and D and E proved too much for her to cope with. The majority of her errors were ABD and ABCE. Here there were too many degrees of freedom to support a high level of performance using such a "gambling" strategy.

Despite the failure of most subjects, there were three subjects in the present study who did acquire the full ABCDE set for both the size and colour series. Were these subjects making a qualitative distinction between size series and colour series? If this was the case, increasing the set size would reveal quantitative differences in performance in terms of both item capacity and fewer trials required to reach criterion. If subjects used the dimension of size to control their search through a monotonic series, they could rapidly control large sets without over-taxing memory. This is because at any point in a monotonic series subjects need only keep track of their current position and the direction of search. At any point the series can be regenerated "on-line." There is no need to recall each item from memory. With a series of unrelated items however, there is no way to generate the next item "on-line". Because each item has to be recalled from memory, there is a limit on the length of strings subjects can control. If subjects failed to take advantage of the fact that the monotonic size series is non-arbitrary (i.e. items are related ordinally along the dimension of size), then they would have to rely on serial control alone to report the series. In this case, the same limitations on set sizes found with arbitrary series would come into effect. Unlike the adult, who would find reporting the monotonic size series easier, the three-year old finds both series equally difficult.

The three subjects in question however did not show set size *restrictions* at five items, rather they had reached the maximum set allowable in this study for both the size and colour series. In this sense, they met the restrictions of the study, rather than their own restrictions in ability. Because of the failure-based design of this study, having not come across set size limitations in these subjects we are limited in what we can say about the source of their success. In particular, there is ambiguity as to whether or not these subjects were using means to control the monotonic size series which were different from those used to control the arbitrary colour series. In Chapter 5, the question of whether or not subjects were making a qualitative distinction between these two strings is addressed.

Chapter 6: Serial control and dimensional control in seriation tasks

Rationale

In the previous chapter we saw that for many young subjects there was a limitation of serial control which resulted in failures at small set sizes in both arbitrary and non-arbitrary serial tasks. However there were three subjects who did not show this equality of failure. Instead, they showed an equality of success at the maximum five-item set for both series. Therefore the question remains of whether or not they made a qualitative distinction between these two series and whether or not they used different means to control them. Subjects taking advantage of ordinal information to constrain their search through monotonic size series could control larger set sizes than by using serial control alone. In this case one would expect to find selective performance in terms of set size in favour of the monotonic size series. In this chapter we use further procedures designed to reveal selective performance.

The most judicious next step would be to increase sets for both series and follow the same failure based analysis. It may be the case that five items was not a large enough set size to bring out a qualitative distinction in means of controlling arbitrary and non-arbitrary series. With small sets there may be too little strain on memory. It may be only with strain on memory that subjects use a strategy to reduce that strain. Although possible in principle, owing to restrictions in available technology at the time of the implementation of this study, it was not possible for us to increase the number of items beyond five.

However, there is already some evidence from Experiments 3.1 & 3.2 that supports the proposition that subjects were not taking advantage of ordinal information to constrain their search through monotonic size series. Firstly, in the over-training phase of Experiment 3.1, subjects had difficulty reporting two size sets which expressed the same ordinal relationship. If subjects were using the monotonic size relationship to constrain search, the absolute values of items should not have produced

performance deficits. Secondly, the poor performance of subjects on the size series delayed response task indicates that they were unable to regenerate the size series on-line from a location other than the first position. Thirdly, subjects required *fewer* trials to learn the ABCDE colour series than they did to learn the ABCDE monotonic size series. If subjects were relying on serial control alone to report the series, they were likely to have found coloured squares easier to identify and distinguish one from the other than sized squares. However it is also possible that subjects were more familiar with the apparatus and the paradigm in colour series training than in size series training. The reduced number of trials to criterion could have been due to an order effect.

However, these conclusions are based on the necessarily consecutive administration of size and colour series training. In order to obtain a stronger measure of whether or not subjects show selective performance on either size or colour series it was necessary to administer both series concurrently. If subjects show selective performance in favour of the monotonic series, then we may provisionally conclude that are using dimensional control. If subjects show performance in favour of the colour series we may conclude that they are using serial control alone. The colour series would be favoured due to the salience of the coloured items as opposed to sized items which only differ slightly one from the other. If subjects again show equality of success, further tests are required. In addition to subjects performing at the maximum allowable set size for both series, two subjects performing at smaller set sizes were included in order to confirm the similarity of their limitations on both strings.

Another means of addressing the question of whether or not a qualitative distinction was being made between the size and colour series by testing subjects on the reverses of both series. This also addresses the issue of the transfer value of subjects' means of controlling a series. If subjects were making a qualitative distinction between size and colour series they would find it easier to reverse the size series rather than the colour series. Reversing a monotonic size series only requires subjects to change the starting pole and direction of search. Reversing an arbitrary colour series is a more difficult operation. Arbitrary series held together purely by the temporal cement imposed by an external agent tend to resist reversal. Subjects would have to transpose the series onto a spatial vector for

reversal or avoid the issue of reversal altogether by relearning it as if it were a novel series.

If subjects are taking advantage of ordinal information to constrain their search through monotonic size series, we would expect that transfer between a size series and its reversal would involve minimal error. When error does occur it should be largely restricted to polar cross-over, (i.e. a choice bias to E in the first position). Transfer between a colour series and its reversal would be less readily achieved, because there is no inherent ordering in the stimuli, and thus no obvious reverse. The series would be learnt in much the same way as a novel sequence of items. The usual distribution of error across internal items should be observed. If subjects are *not* taking advantage of ordinal information to constrain their search through the size series we would expect that transfer in both cases would be difficult with subjects essentially having to learn both "reverse" series as if they were novel series.

The following experiments have two main goals. Firstly, we investigate young subjects' concurrent performance on non-arbitrary, monotonic size series and arbitrary colour series. Secondly, we assess the transfer value of subjects' means of controlling series, with respect to reversed series. In both instances we seek to discover whether or not there is selective performance in favour of the monotonic size series which indicate the use of dimensional control over serial control.

EXPERIMENT 4.1

Subjects

The same children as in Experiment 3.2, with the exception of Cath, Nicole, and Andrew, served as subjects. Cath and Nicole were dropped from the study at this point because their productions were severely limited (to three items or less). There would be little to gain in investigating concurrent performance. Andrew was treated separately due to his high level of performance in training on the monotonic size series. His performance is taken up in Experiment 4.2 below.

Apparatus & Stimuli

The apparatus was the same as used in Experiments 3.1 and 3.2. For size series tasks the stimuli were the "training set" in Experiment 3.1. As before, stimuli were colour coded for condition: red for monotonic descending series and green for monotonic ascending series. For colour series tasks the stimuli were the same as in Experiment 3.2.

Design & Procedure

The design of Experiment 4.1 is shown in Figure 6.1. Subjects received size and colour series tasks in alternating daily sessions. Subjects' first and third sessions were the size series task and the second and fourth sessions were the colour series task. All subjects maintained their original size and colour series assignments.

Daily Trial Blocks Test (Size v. Colour). The procedure was identical to the usual training procedure except that subjects were immediately confronted with full ABCDE set and required to find the correct path through the set. Commission of four successive errors resulted in the reduction of the set by one item (e.g., ABCDE to ABCD). Six successive correct responses were required to add a further item (ABCDE). The session was terminated when the subject reached a criterion of six successive correct responses on the ABCDE series or after completion of no more than twenty trials (including correction trials). This number of trials was chosen based on the average number of trials per session tolerated by subjects in Experiment 3.2. Sessions were terminated if subjects showed signs of tiring.

If after one session of each series subjects did not reach criterion, subjects were returned to the original training procedure (beginning with one item, then two, etc.) for two further sessions, one of each series.

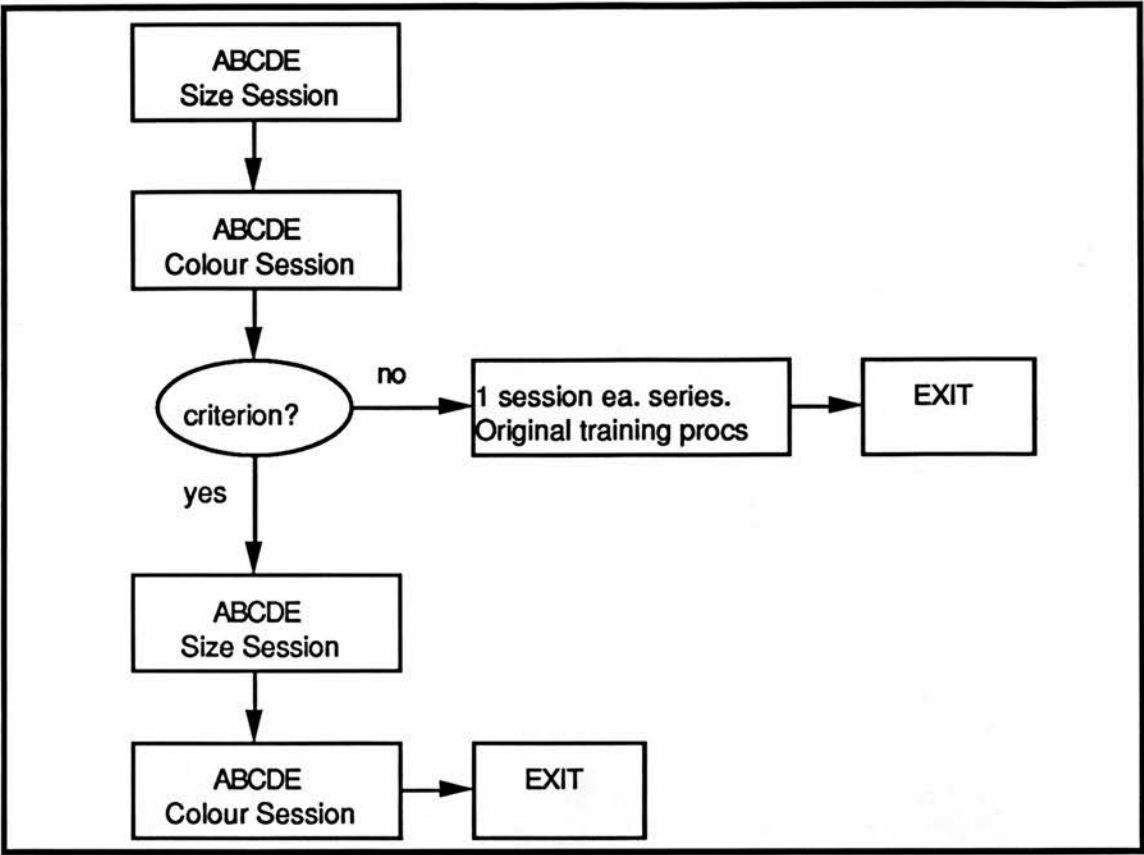


Figure 6.1: Experiment 4.1 Design

Results

Performance after the first two sessions

Performance is summarised in Table 6.1. There was little difference in performance between size strings and colour strings, however there was a striking difference in performance among subjects.

SUBJECT	SERIES	TRIALS	OUT COME
ALEXA	size	1 1	ABCDE crit
	colour	1 4	ABCDE crit
	size	8	ABCDE crit
	colour	1 1	ABCDE crit
YVONNE	size	6	ABCDE crit
	colour	7	ABCDE crit
	size	6	ABCDE crit
	colour	6	ABCDE crit
GILL	size	3	ABCD exit
	colour	9	ABC exit
	size*	2 3	ABCD exit
	colour*	1 9	ABCD exit
ZACH	size	7	ABC exit
	colour	3	ABCD exit
	size*	1 3	ABCD exit
	colour*	2 0	ABC crit

Table 6.1: Trial Blocks Test performance

Both Gill and Zach were achieving a low level of performance on both tasks. Gill immediately “crashed” from ABCDE to ABCD on the size series task and refused to carry on after just two ABCD trials. In the colour series task she rapidly (after only two trials at ABCDE and ABCD) “crashed” from ABCDE to ABCD to ABC. Her session was terminated after she completed twenty trials (including correction trials). Similarly, Zach immediately “crashed” from ABCDE to ABCD on the size series task, and after 5 trials (three contained multiple errors) with the ABCD series “crashed” to ABC. In the colour series task Zach immediately “crashed” from ABCDE to ABCD and refused to carry on after two ABCD trials, both of which contained multiple errors.

Alexa and Yvonne, on the other hand, were achieving a high level of performance on both tasks. Alexa required roughly the same number of trials to achieve criterion on both the size and colour task. However, on the colour series she immediately “crashed” to ABCD, but then returned to ABCDE with only one error. She then reached criterion on ABCDE in the minimum number of trials. Yvonne reached the ABCDE size series

criterion in the minimum number of trials and the ABCDE colour series criterion with only one error.

Performance after the second two sessions

Due to their poor performance in the first two sessions, Gill and Zach received two further sessions of each series, but under the original training procedure. Alexa and Yvonne carried on for two further sessions under the daily trials blocks test procedure.

Gill and Zach continued to show poor performance (see Table 6.1). Both subjects reached the AB and ABC size criteria with the minimum number of trials but failed to reach criterion on the ABCD series. In the colour task, Gill required the minimum number of trial to reach the AB criterion and reached the ABC criterion with only one error. However she was unable to reach the ABCD criterion and exited after six trials. Zach reached the AB criterion in 8 trials and the ABC criterion in 12 trials. He refused to work on the ABCD series claiming it was "too hard" for him.

Alexa and Yvonne continued to show high levels of performance (see Table 6.1) on both size and colour series. Alexa required 8 trials to reach criterion on the size series and 11 trials to reach criterion on the colour series. Yvonne required the minimum number of trials to reach criterion on both series.

Error analysis for Gill and Zach.

The majority of errors (all errors for both series) were forward errors of the one-step variety (65%) as has been seen in previous training and testing. There were also a substantial number of reference errors (29%). In the ABCD trials, which represented the limit of subjects' abilities, errors were distributed 33% at A, 33% at B, and 33% at C for size series, 0% at A, 60% at B, and 40% at C for colour series. Errors were distributed 18% at A, 45% at B, and 36% at C overall.

RT analysis for Alexa and Yvonne

Inter-item response time data were taken from all correct ABCDE trials for both sessions of size and colour series. The usual procedures for calculation and "clean up" were applied (see Experiment 2). Average

decision times for completion of ABCDE series and inter-item RT can be found in Table 6.2.

Subject	ser.	A	B	C	D	E	Total
Alexa	size	200 (25)	190 (23)	166 (20)	153 (19)	103 (13)	811
Alexa	col	224 (23)	261 (27)	236 (24)	146 (15)	102 (11)	970
Yvonne	size	334 (38)	198 (23)	114 (13)	139 (16)	82 (9)	868
Yvonne	col	406 (33)	290 (24)	177 (15)	209 (17)	133 (11)	1214

Table 6.2: Mean decision time and IRT (%of decision time in parenthesis)

Decision times for the size series were shorter than for the colour series. Alexa took on average 8.11 seconds to complete a size series and 9.70 seconds to complete a colour series, a difference of 1.59 seconds. Yvonne took 8.68 seconds to complete a size series and 12.14 seconds to complete a colour series, a difference of 3.46 seconds.

For Alexa the pattern of inter-item RT was different for size and colour series (see Figure 6.2). For the size series she showed the same descending pattern as in size series over-training. In the colour series the pattern of descent was similar, however it was the second item which required the greatest amount of time. This was different to the pattern of inter-item RT in colour series over-training where the first three items had very similar latencies with less time required for the last two items.

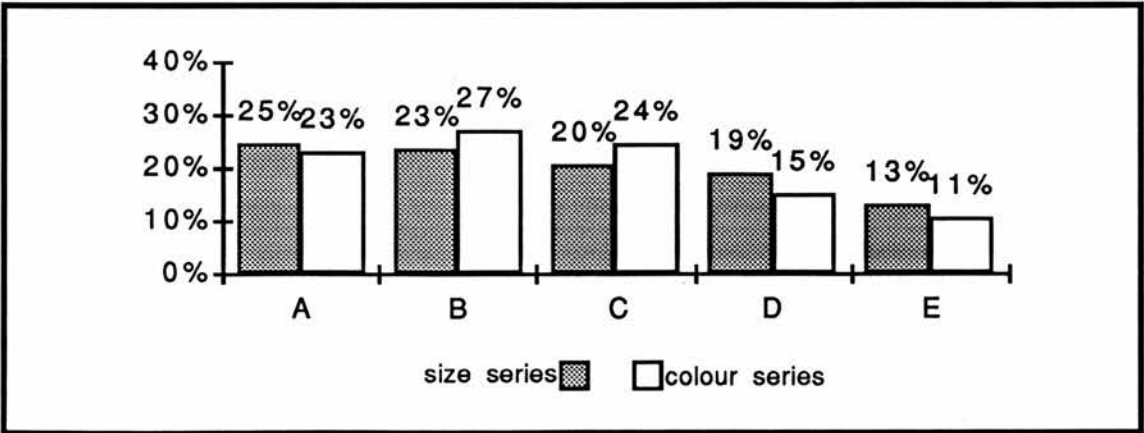


Figure 6.2: Size v. colour series IRT (Alexa)

For Yvonne the pattern of inter-item RT was similar for size and colour series (see Figure 6.3). Her pattern suggests that she may have been spontaneously reporting both ABCDE series in two phrases: ABC-DE. In over-training, she showed this pattern with size series, but not with colour series. A feature common to all Yvonne's series was that the first item requires the most decision time.

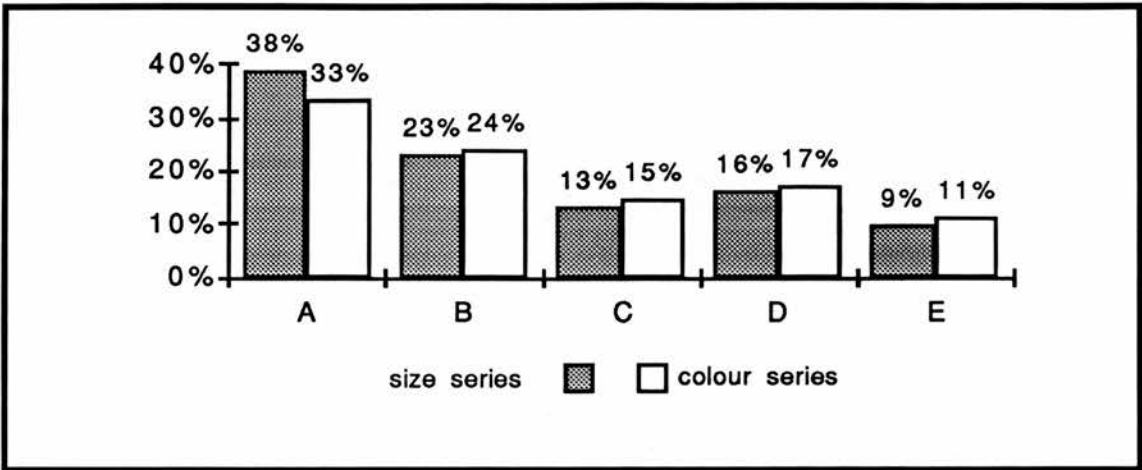


Figure 6.3: Size v. colour series IRT (Yvonne)

Discussion of Experiment 4.1

The acquisition phase finding that subjects were equally item limited on both size and colour series was confirmed. All subjects held the same level of performance on size and colour series in terms of item limitation. Two subjects were severely limited to three items on both series, whilst two other subjects could both control the full five-item set. Two of our subjects were clearly limited to three-item series regardless of whether it was a size or colour series. By testing subject performance both from five items down and from one item up, we converged on an item limitation in terms of both the minimum and maximum number of items subjects could control. For these young subjects the ordinal nature of the monotonic size set held no advantage over the arbitrary colour string. From this finding, together with error information, it may be interpreted that subjects used the same means to report both series. Owing to the severe item limitation and the ordinal position of errors in ABCD trials, this means is likely to be a simplistic exclusion strategy rather than serial control.

It is unclear whether subjects who could control five items for both size and colour series were making a qualitative distinction between the two series, because they were operating at the ceiling with respect to the limits of this procedure. For these subjects five items was not enough to bring out possible differences in means of controlling size and colour strings. Although Alexa required more trials to reach criterion with the colour series the difference was very small (only 3 trials). For Yvonne there was virtually no difference at all. The only clear difference between series was in terms of decision time. The colour series required more time to complete than the size series. However, there was no evidence that faster decision times on size series were due to subjects taking advantage of ordinal information. Yvonne was spontaneously reporting both series in phrases. Phrasing reduces memory load for arbitrary strings by creating larger building blocks which are less likely to migrate out of order and lead to mistakes (McGonigle, 1987). In highly routinized performance, that Yvonne was phrasing both strings may indicate that she treated both series as arbitrary strings.

EXPERIMENT 4.2

Subjects

Alexa, Yvonne, and Andrew.

Apparatus & Stimuli

The apparatus was the same as used above. For size series tasks the stimuli were the "training set" in Experiment 3.1. As before, stimuli were colour coded for condition: red for monotonic descending series and green for monotonic ascending series. For colour series tasks the stimuli were as in Experiment 3.2. In reverse colour series tasks the stimuli were identical to ordinary colour series stimuli except that they had a small (0.5 cm) white dot in the centre to distinguish them.

Design & Procedure

Andrew

Because of this subject's success in maintaining a high level of routinized performance in Experiment 3.1, he was singled out for immediate inverse

series training and testing. After completing the delayed response tests and restating the routinization criterion in Experiment 3.1, the subject was trained on the "inverse" of the monotonic descending series (i.e. monotonic ascending). After meeting the routinization criterion in Experiment 3.2, the subject was trained on the "inverse" of his assigned colour series. His Experiment 3.2 colour series was red, green, yellow, blue, pink, thus the inverse was pink, blue, yellow, green, red.

Training & Over-training. The subject followed the training procedure as stated in Experiment 3.1 & 3.2. The following additional instructions were given, "Today we have a different game. Look at the squares carefully and give this game a try." Immediately after the subject reached the six successive correct trials criterion with the ABCDE series he was asked the verbal probe questions appropriate to the size or colour series (see Experiments 3.1 & 3.2). Over-training proceeded as previously, however for the size series only the training stimulus set was used. The decision was taken to eliminate the test set of stimuli from over-training in this case so that the effect of inverting the series would not be confounded with the difficulties of coping with two stimulus sets. The subject was required to reach a criterion of nine correct trials out of ten in order to proceed to the test phase.

Daily Trial Blocks Test. The subject followed the "daily trials blocks" test procedure as stated in Experiment 4.1. The first and third sessions were the subject's original training direction (descending size, RGYBP colour). The second and fourth sessions were the inverse direction (ascending size, PBYGR colour). Commission of four successive errors resulted in the reduction of the set by one item (e.g., ABCDE to ABCD). Six successive correct responses were required to add a further item (ABCDE). The session was terminated when the subject reached six successive correct responses on the ABCDE series or after completion of no more than twenty trials (including correction trials).

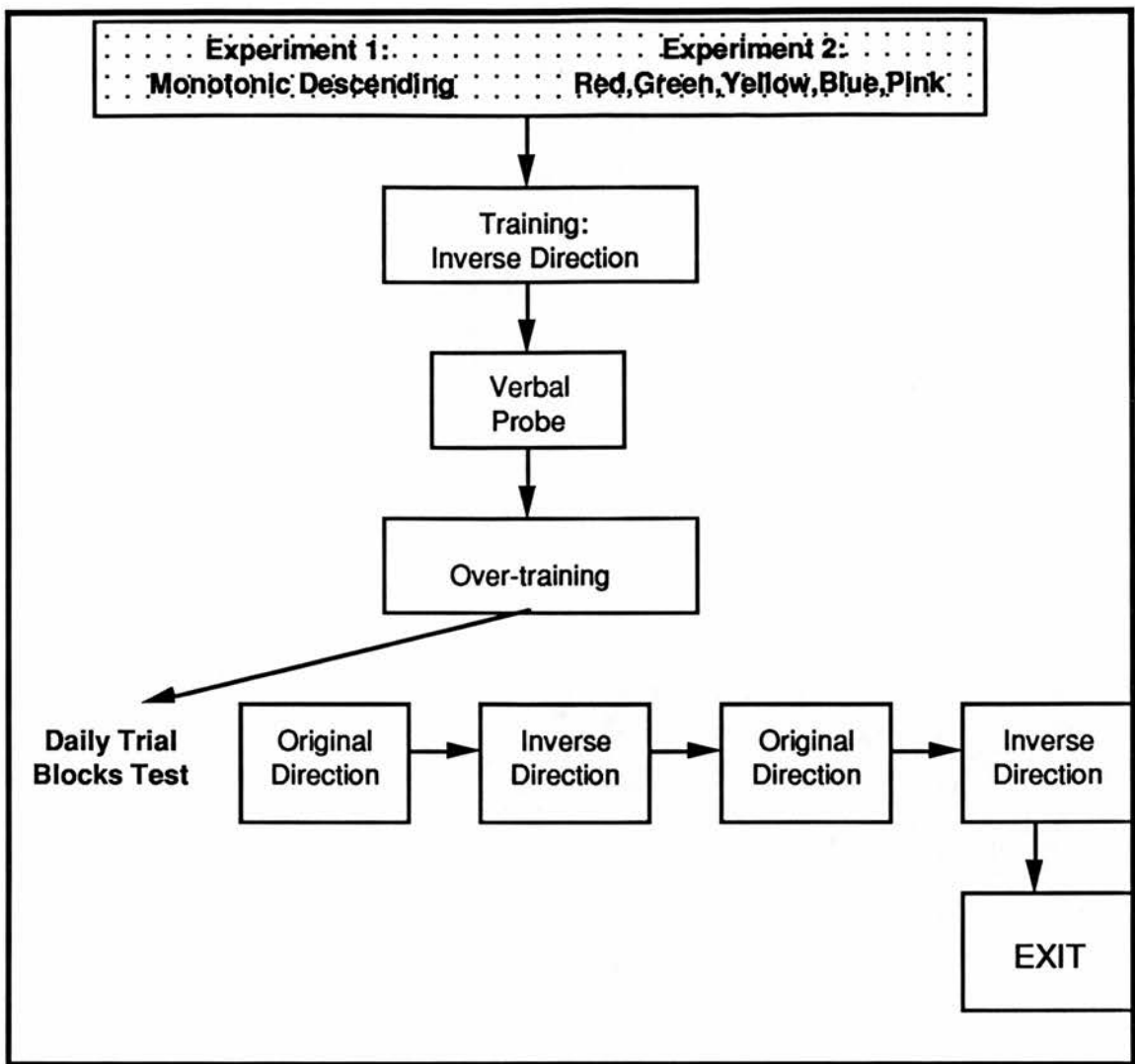


Figure 6.4: Experiment 4.2 design (Andrew)

Alexa and Yvonne

Subjects were tested on the inverses of both size and colour series. Alexa was tested on the inverse size series first and the inverse colour series second. Yvonne was tested on the inverse colour series first and the inverse size series second.

Refreshment. Before beginning the inverse series tests, subjects received one session of over-training trials in order to refresh performance. Subjects were required to reach a criterion of six successive correct trials in no more than two sessions or they were removed from the study.

Transfer to Inverse Series. The procedure was identical to the usual training procedure except that subjects were immediately confronted with full ABCDE set and required to find the correct monotonic path through the set. The decision was taken to test inverse series transfer using the full ABCDE set immediately (rather than using the parsed set training procedure) in order to avoid the effect of subjects learning a new series which they might see as entirely divorced from its "inverse." Commission of six successive errors resulted in the reduction of the set by one item (e.g. ABCDE to ABCD). This "error criterion" was increased from four to six in order to provide a larger window on error information. As before, six successive correct responses were required to add a further item (ABCDE). The session was terminated when the subject reached six successive correct trials on the ABCDE series or after completion of no more than twenty trials (including correction trials). Subjects were required to reach the six successive correct trials criterion for two successive sessions.

Return to Original Series. Subjects followed the same procedure as above, but using their original training direction. Subjects were required to reach a criterion of six successive correct trials.

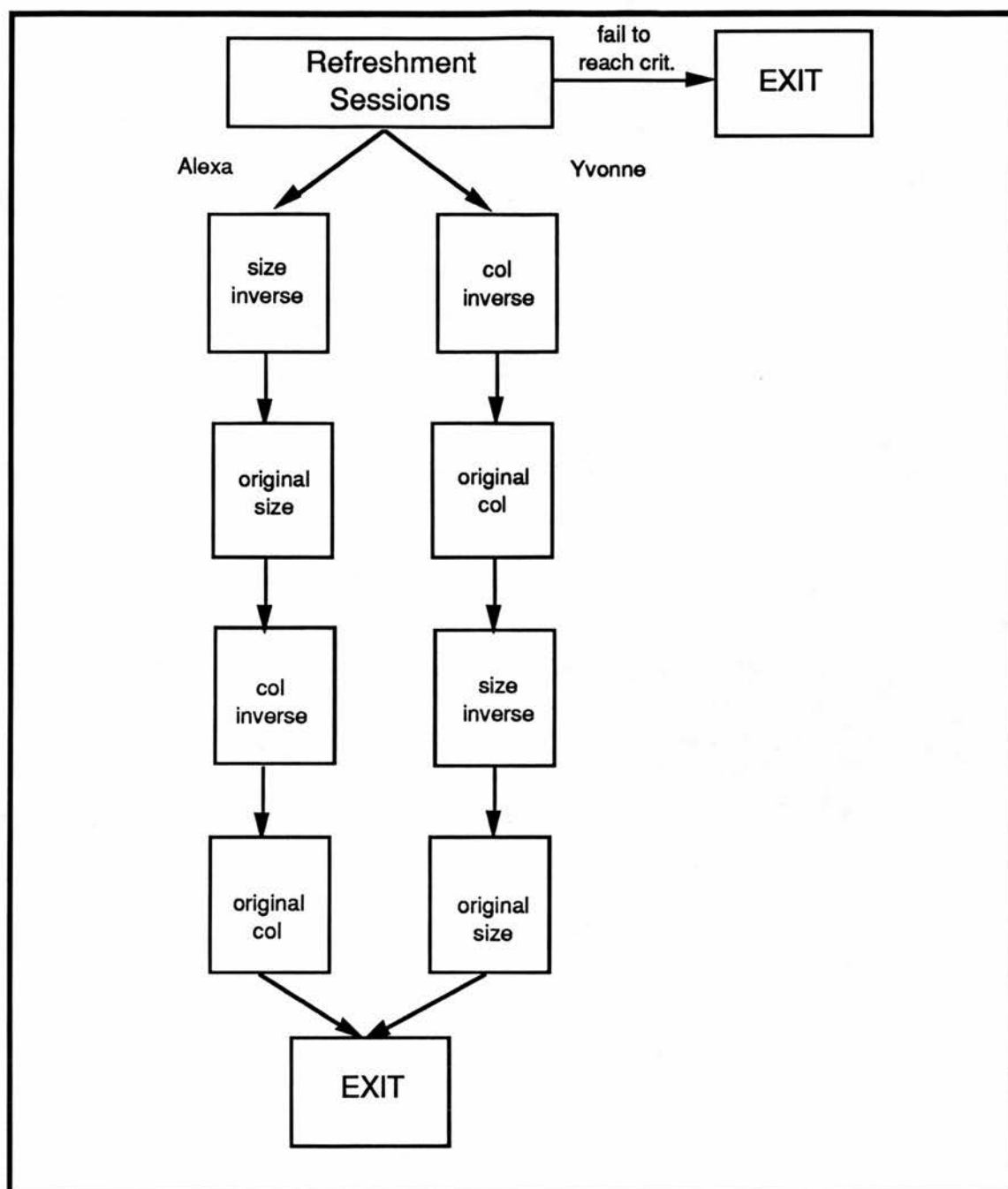


Figure 6.5: Experiment 4.2 design (Alexa & Yvonne)

Results (Andrew)

Reverse Size Series Acquisition

Trials to criteria. The subject required 52 trials to reach the ABCDE criterion for the monotonic ascending series. This was precisely the same number of trials he required to reach criterion with his first size series,

(monotonic descending). The number of trials required to reach each criterion were 6 (the minimum) for AB, 14 for ABC, 26 for ABCD, and 6 for ABCDE. For the monotonic ascending series the subject had the greatest difficulty acquiring the ABCD series; he spent 50% of total trials reaching criterion on the ABCD series. Another site of difficulty was the ABC series, where 27% of total trials were required to reach criterion. This pattern of distribution of trials to criterion was different to the pattern for acquisition of the first series, where ABC required the most trials.

Error. The subject made 11 errors before reaching the ABCDE criterion. All ABC training errors were reference errors. In ABCD training all errors were of the one-step forward variety. 71% of these were AC errors and 29% were ABD errors. Therefore, the majority of errors occurred at the second position (B). There were no errors on ABCDE training. The error picture was similar for the subject's first size series where 13 errors were made; 83% of errors were of the forward one-step variety and 58% of errors occurred at the second position (B).

Verbal Probe

Consistent with prior performance, the subject's verbal command of the series was not good. He was able to express that "the little one" was first and "the big one" was last. When asked questions about which block comes next or before he simply alternated between "big" and "little."

Over-training

Performance. The subject required 3 sessions (65 trials) to reach the 9 correct trials out of 10 criterion. His performance was quite high on a per session basis --- 74%, 74%, and 82% correct respectively (75% overall). The subject made 16 errors, 94% of which were forward errors. Most of these were of the one-step variety (67%), however there were a small proportion of two-step (AD and ABE) and four-step (AE) forward errors as well, 20% and 13% respectively. There was one reference error in which the choice was E, the opposite pole. The majority of errors (63%) occurred at the second position (B). The other main site of error was C where 25% of errors occurred.

Decision Time and Inter-Item RT. Average decision time for the completion of the monotonic ascending series during over-training was 4.58 seconds. This compares with 6.70 seconds for the monotonic descending (training set only). Considering the similarity of performance in terms of trials and errors to criterion, it was likely that this faster time reflects the subject's increased familiarity with the apparatus and stimuli. The mean inter-item RT was 1.32, 1.04, 0.76, 0.76, and 0.70 seconds respectively. Expressed as a percentage of total decision time RT distributes 29%, 23%, 17%, 17%, and 15% respectively. Figure 6.6 compares inter-item RT for both size series. The overall pattern was similar, however in the second size series, the subject required a greater percentage of time to report the first item.

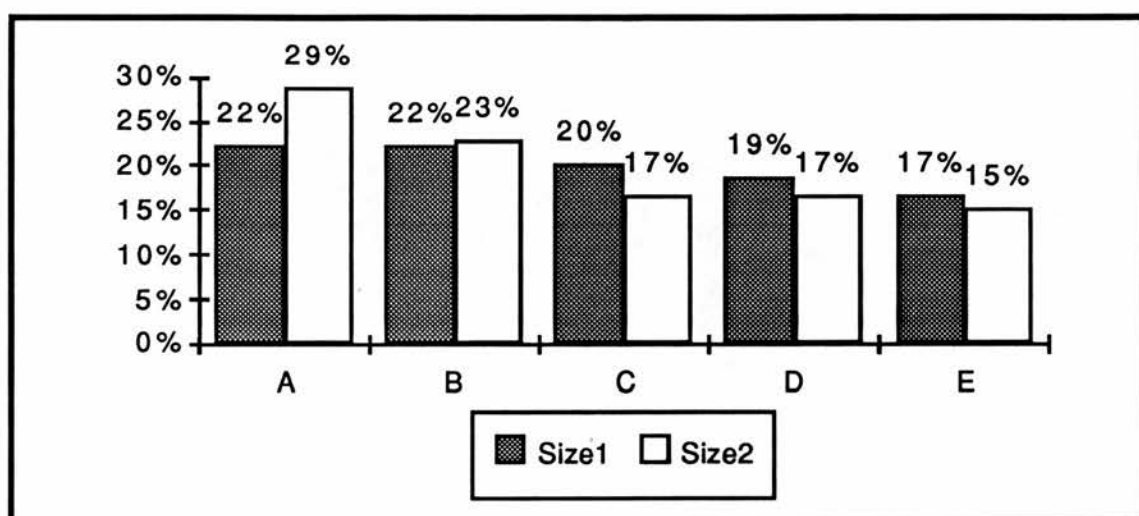


Figure 6.6: Comparison of IRT for 2 size series (Andrew).

Trial Blocks Test

Trials to criteria. Performance is summarised in Table 6.3. There was a clear difference between performance on the two size sets, with the subject showing better performance on the first series (monotonic descending). Although he was able to reach the ABCDE criterion on the first size series in both sessions, he required 23 trials in the first session and 17 trials in the second session. Further, in the first session he immediately made 8 consecutive errors, "crashing down" to the ABCD series and then the ABC series. After reaching the ABC criterion in 9 trials he was able to reach the ABCD and ABCDE criteria in the minimum number of trials. In the

second size 1 session the subject again immediately “crashed down” to the ABCD series and then required 7 trials to reach the ABCD criterion and 9 trials to reach the ABCDE criterion.

In both sessions of testing on the second size series (monotonic ascending) the subject was unable to reach the ABCDE criterion. Neither was he able to reach the ABCD criterion after “crash back.” In the first session he immediately made 12 consecutive errors resulting in reduction from ABCDE to ABCD to ABC to AB. After reaching the AB criterion in 6 trials and the ABC criterion in 8 trials he was unable to reach the ABCD criterion after 5 trials and the session was terminated after a total of 22 trials. In the second size 2 session the subject immediately made 4 consecutive errors resulting in reduction from ABCDE to ABCD. After only one correct trial, he made 4 consecutive errors and “crashed down” to ABC. After reaching the ABC criterion in 6 trials he was unable to reach the ABCD criterion after 7 trials and the session was terminated after a total of 17 trials.

SUBJECT	SERIES	TRIALS	OUTCOME
ANDREW	size 1	23	ABCDE crit
	size 2	22	ABCD exit
	size 1	17	ABCDE crit
	size 2	17	ABCD exit

Table 6.3: Trial blocks test performance, size series (Andrew)

Error. The subject was 80% correct on size 1 trials versus 67% correct on size 2 trials. For both size series the main source of error was failure to correctly report the first item, accounting for 50% of error in size 1 trials and 62% of error in size 2 trials. All other errors were the typical AC and ABD one-step forward errors.

Error Choice. In the first ABCDE trial of the first size 1 session the display was C, D, A, B, E. The subject’s choice was E. In the correction trials he chose B (adjacent in the display), then C (at the other extreme of the display), and finally D (adjacent to C). In the first ABCDE trial of the second size 1 session, the display was C, E, D, B, A. The subject’s choice was D. In the correction trials he chose D again, then C, and finally made an ABD error.

In the first ABCDE trial of the first size 2 session the display was A, D, C, E, B. The subject's choice was E. In the correction trials he chose C (adjacent to E), then D (adjacent to C), and finally B (at the other extreme of the display). In the first ABCDE trial of the second size 2 session, the display was A, E, B, C, D. The subject's choice was E. In the correction trials he chose D (at the extreme right), then D again, and finally chose C (adjacent to D).

Reverse Colour Series Acquisition

Trials to criteria. The subject required 57 trials to reach the full five-item ABCDE criterion for the "reverse" colour series, whereas he required 38 trials to reach the full five-item ABCDE criterion on the first colour series. The number of trials required to reach each criterion were 6 (the minimum) for AB, 13 for ABC, 17 for ABCD, and 21 for ABCDE. The subject required steadily more trials to reach each criterion. Expressed as a percent of total trials to criterion, the subject required 11%, 23%, 30%, and 37% of trials respectively. This pattern of distribution of trials to criterion was different to the pattern for acquisition of the first colour series, where ABC required the greatest percentage of trials.

Error. The subject made 13 errors before reaching the ABCDE criterion. Overall error distributed 62% forward errors 8% backward errors and 31% reference errors. All ABC training errors were reference errors to B. In ABCD training all errors were one-step forward errors (two AC and one ABD), with the exception of one backward error. All ABCDE training errors were one-step forward errors except for one reference error to B. These were three AC errors, two ABD errors and one ABCE error. Therefore the majority of ABCDE errors were at the second position (B). Although fewer errors were made in training on the first series (8 versus 13), the training error picture in terms of error type was similar to the subject's first colour series. In colour 1 training 50% of errors were forward errors, 13% were backward errors, and 38% were reference errors.

Verbal Probe

Consistent with prior performance, the subject's verbal command of the series was not good. When asked to freely recall the colours he reported

only pink, green, yellow, blue (A, D, C, B). When prompted for the colours from first to last, he reported red first (E), then blue (B), and repeated red and blue again and then said he didn't know. When prompted for the colours from last to first he reported red, blue, green, pink (E, B, D, A) and said that was all. He was also incorrect on both the gap-filling and exclusion questions, responding "blue and yellow" (B, C) to both.

Over-Training

Performance. The subject required 1 session (13 trials) to reach the 9 correct trials out of 10 criterion. Overall he was 85% correct. He made only two errors --- one forward error (ABD) and one backward error (ABCDB).

Decision Time and inter-item RT. Average decision time for the completion of the reverse colour series during over-training was 7.96 seconds. This compares with 11.34 seconds for the first colour series. Again, it is likely that less time was required to report the second colour series due to the subject's increased familiarity with the apparatus and stimuli. The mean inter-item RT was 2.11, 2.03, 1.93, 0.80, 1.08 seconds respectively. Expressed as a percentage of total decision time RT distributes 27%, 26%, 24%, 10%, and 14% respectively. Figure 6.7 compares inter-item RT for both colour series. The overall pattern was similar except for the reversed pattern of latencies to C and D. Also there was no pattern of phrasing found for the second colour series.

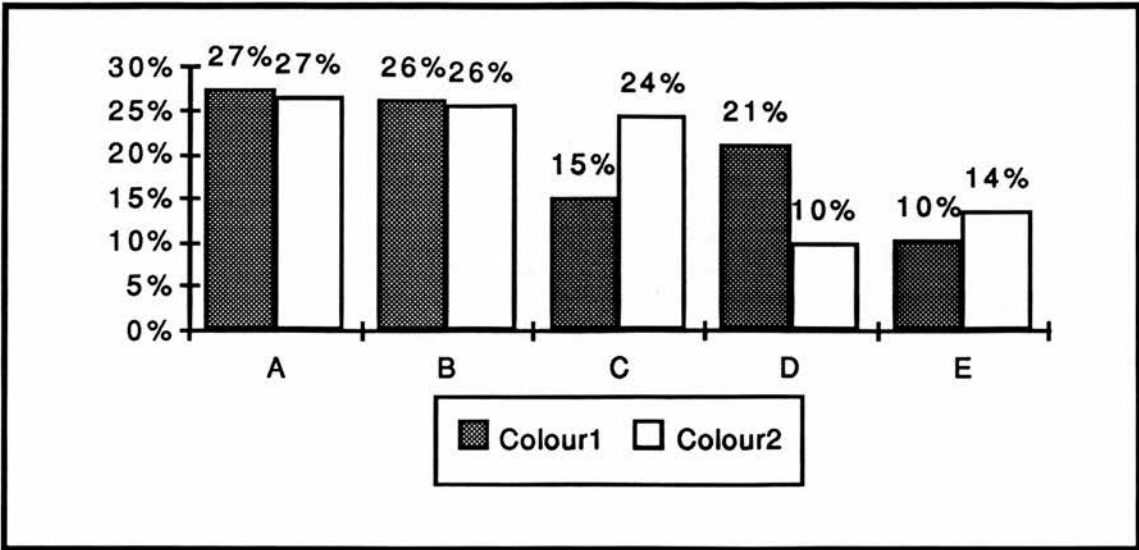


Figure 6.7: Comparison of IRT for 2 colour series (Andrew).

Trial Blocks Test

Trials to criterion. Performance is summarised in Table 6.4. There was a clear difference between performance on the two colour series, with the subject showing better performance on the second colour series. In the first session of the first colour series, the subject immediately made 8 consecutive errors resulting in reduction from ABCDE to ABCD to ABC. The subject reached the ABC criterion in 12 trials, and after one ABCD trial the session was terminated.

In the second session, the subject immediately made 4 consecutive errors resulting in reduction from ABCDE to ABCD. He reached the ABCD criterion in the minimum number of trials, but was unable to reach the ABCDE criterion after 10 trials. In both sessions of the second (reverse) colour series the subject required 18 trials to reach the ABCDE criterion. In the first session these 18 trials included a “crash down” to the ABCD series.

SUBJECT	SERIES	TRIALS	OUT COME
ANDREW	col 1	15	ABCD exit
	col 2	18	ABCDE crit
	col 1	17	ABCDE exit
	col 2	18	ABCDE crit

Table 6.4: Trial blocks test performance, colour series (Andrew)

Errors and Error Choice. The subject was 72% correct on colour 1 trials and 81% correct on colour 2 trials. For both colour series the main source of error was the tendency to skip one item forward in the series, accounting for 67% of error in colour 1 trials and 43% of error in colour 2 trials. It is interesting to note error choice in the first trials (and correction trials) of colour 1 sessions in which the subject immediately “crashes down” to ABCD. In the first ABCDE trial of the first colour 1 session, the subject’s choice was E. In the correction trials he made AE, AD, and AC errors. In the first ABCDE trial of the second colour 1 session, the subject’s choice was D. In the correction trials he made AD, AC, and ABCE error.

Results (Alexa & Yvonne)

Refreshment

Both subjects were able to meet the 6 correct trials in a row criterion for both series in one session each. Size series criteria were met by both subjects in the minimum number of trials, whereas for the colour series Alexa required 7 trials and Yvonne required 14 trials to reach the criterion.

Acquisition of Reverse Series

Trials to criterion. Both subjects were able to reach the ABCDE criterion for both “reverse” series. For the size 2 series Alexa required 10 trials and Yvonne required 9 trials to reach criterion. Neither subject required reduction to the ABCD series. Both subjects required more trials to reach the colour 2 series criterion than the size 2 criterion, nearly three times more in Yvonne’s case. Alexa required 16 trials and Yvonne required 25 trials to reach criterion. Both subjects required reduction to the ABCD series.

Error Choice. In the first ABCDE trial of the size 2 acquisition session, Alexa’s choice was E. In the correction trials she chose D (second biggest) and then made an AE error. Yvonne’s also chose E. In the correction trials she chose D, then C, then B and then made an ABCE error.

In the first ABCDE trial of the colour 2 acquisition session, Alexa’s choice was E. In the correction trials she chose D, then B, then E again, then C

and finally made an ABD error. Yvonne's also chose E. In the correction trials she made the errors AE, AC, ABE, ABD, and ABE.

Trial Blocks Test

Performance. After each "reverse" acquisition session the subjects had one further session of over-training to assess reliability. In the following session they were again confronted with the original series. Table 6.5 shows the number of trials required to reach criterion for each series. The bottom panel shows percent correct.

Series	Alexa	Yvonne
size 2 acq	10	9
size 2	10	8
size 1	6	7
col 2 acq	16	25
col 2	11	6
col 1	7	9

Series	Alexa	Yvonne
size 2 acq	70%	67%
size 2	90%	63%
size 1	100%	86%
col 2 acq	75%	76%
col 2	73%	100%
col 1	86%	89%

Table 6.5: Experiment 4.2 performance

In "reverse" size series over-training sessions both subjects showed performance similar to the acquisition session, although Alexa showed a considerable improvement in errors (if not trials) to criterion. Alexa's only error was a typical one-step forward (AC) error. Yvonne's first error was an incorrect first choice to E. Her two subsequent errors were both typical one-step forward (ABD) errors. Subjects easily transferred back to their original size series, reaching criterion in the minimum, or near

minimum in Yvonne's case, number of trials. The only error was a typical one-step forward (AC) error.

In "reverse" colour series over-training sessions Yvonne showed improvement in performance over the acquisition session, whereas Alexa showed similar performance to the acquisition session. Alexa made three errors two AC errors and one AE error. Subjects easily transferred back to their original colour series, reaching criterion with only one error each. Alexa made a typical one-step forward (AC) error. Yvonne made an ABA backward error, which was caused by the subject interrupting her report of the series to make a comment to the experimenter. When she returned to the series, she began again at A.

Decision Time and Inter-Item RT. Inter-item response time data were taken from all correct ABCDE trials in the over-training sessions for all four series. The usual procedures for calculation and "clean up" were applied. Average decision times and inter-item RT for all series can be found in Table 6.6 (upper panel for Alexa, lower panel for Yvonne).

Alexa

Series	A	B	C	D	E	Total
Size1	429 (38)	234 (21)	242 (22)	123 (11)	96 (9)	1124
Size2	359 (29)	326 (27)	182 (15)	229 (19)	136 (11)	1231
Colour1	1180 (50)	494 (21)	288 (12)	293 (12)	108 (5)	2363
Colour2	503 (34)	296 (20)	353 (24)	207 (14)	133 (9)	1491

Yvonne

Series	A	B	C	D	E	Total
Size1	489 (41)	188 (16)	247 (21)	170 (14)	86 (7)	1180
Size2	308 (30)	195 (19)	278 (27)	137 (14)	92 (9)	1010
Colour1	502 (36)	364 (26)	183 (13)	233 (17)	109 (8)	1391
Colour2	473 (33)	254 (18)	303 (21)	209 (14)	204 (14)	1443

Table 6.6: Mean Decision Time and Inter-Item RT (% of Decision Time in parenthesis)

As was found in Experiment 4.1, decision times for size series were shorter than those for colour series. Series of the same type tended to take similar

amounts of time to complete. There was less than 2 seconds difference between decision times, except in the case of Alexa's performance on colour series, where colour 1 took 8.72 seconds longer to report than colour 2. Examination of inter-item RT showed that this was likely to be due to long latencies to the first item. The subject took long "breaks" between trials to make comments to the experimenter.

For Alexa, the pattern of inter-item RT was similar for size 1 and colour 2. Her pattern of latencies suggest that she was spontaneously reporting these two series in two phrases, AB-CDE. For the size 2 and colour 1 series, however, her pattern of latencies suggest that she was phrasing these series as ABC-DE.

For Yvonne, the pattern of inter-item RT was similar for all series (except colour 1). Her pattern of latencies suggests that she was spontaneously reporting series in two phrases, AB-CDE. If the "expense" for inter-item RT is expressed in terms of ranks (1 = least time, 5 = most time), we find that for Yvonne A = 5, B = 3, C = 4, D = 2, E = 1. Analysis using Kendall's coefficient of concordance showed that this ranking ($W = 0.87$) was significant at the $p < 0.01$ level. The AB-CDE pattern of phrasing was also found in the previous study (Experiment 2) of older subjects' reporting of ABCDE colour series.

Results Summary

Andrew acquired the "reverse" series of both his previous size and colour series. He made no savings in terms of trials required to reach the ABCDE criteria. He required the same number of trials to learn the reverse size series as he did the original size series and 1.5 times as many trials to learn the reverse colour series as he did the original colour series. Acquisition errors were similar in all series, typically one-step forward errors at B or C (i.e., AC and ABD errors).

In over-training trials, Andrew was able to reach the 9 correct trials out of 10 criterion in all cases. His pattern of inter-item RT was different for size and colour series. In both size series he showed a steady decline in time required to report an item from A to E. In the colour 1 series he may have been reporting the series in three phrases, A-BC-DE. However in the colour 2 series, he required more or less the same amount of time to

report the first three items with a rapid reduction in time to report the last two items.

In daily trial blocks tests the subject had great difficulty in switching from a series to its reverse. In size series trials blocks he performed better on the first size series, however in colour series trials blocks he performed better on the second colour series. His first error in the first trial of each session was to the opposite pole. He then typically searched adjacent items until he found the correct item.

Alexa and Yvonne followed a different procedure for learning reverse series. They too were able to learn the "reverse" series of their previous size and colour series. Both subjects had greater difficulty learning the colour "reverse" series in terms of trials required to reach criterion and relegation to the ABCD series. The first error in the first trial of acquisition sessions was always to the opposite pole. Thereafter, Alexa typically made forward errors to E (i.e. AE and ABE), whilst Yvonne typically searched items in reverse order (i.e. E, D, C, B) until she found the correct item.

In daily trial blocks tests subjects did not have difficulty in switching from a series to its reverse. Errors followed patterns which could be expected normally. Both subjects were found to be spontaneously reporting all series in phrases. Yvonne used an AB-CDE pattern of phrasing, whilst Alexa used AB-CDE for size 1 and colour 2 series and ABC-DE for size 2 and colour 1 series.

Discussion of Experiment 4.2

Although subjects showed some differences in acquisition and maintenance of size and colour series "reverses," there was no clear evidence that this due to their taking advantage of ordinal information to constrain their search through size series. For Andrew, the first piece of evidence to support a lack of dimensional control is that he showed no savings in terms of trials to criterion to learn the reverse size series. It may be the case that he treated the reverse as if it were simply a novel, arbitrary series. For acquisition of the colour reverse series he required 1.5 more trials than he did to acquire the original colour series. When all four acquisitions are taken together, the number of trials required to learn

the colour reverse series (57) is closer to that required for the size series (52 each) than the original colour series (38). It may simply be the case that some factor particular to the first colour series made it easier for Andrew than other series of any type.¹ That the first colour series was most rapidly learnt may have been due to the subject recruiting phrasing to assist memory. Why he did not “phrase” the second size series, again may be due to a local factor.

Secondly, difficulty in switching between series and their inverses for both size and colour series in daily trials blocks tests further suggests that for Andrew monotonic size series had no special status. In particular, error patterns in first trials may be interpreted as indicating that the subject used simple exhaustive search as a means to discover both size and colour reverse series. Although this pattern of error might be expected for reversing arbitrary series, if the subject were using the size dimension to control his search through size series, he should have been able to predict the first item after one error (by polar cross-over). In fact, Andrew was not successful in switching between two series of either type. One series was always dominant over the other.

Although Alexa and Yvonne had greater difficulty acquiring the colour reverse series it is not clear that this is due to subjects making a qualitative distinction between monotonic size series and arbitrary colour series. The first piece evidence to support a lack of dimensional control is exhaustive search in trials where the subjects were first exposed to reverse series. Yvonne’s error choice data showed exhaustive search in the first trials of both size and colour reverse series, indicating that she was unable to predict the reverse monotonic size series (based on what the first item was not). Alexa’s choice data showed clear exhaustive search on the colour reverse series only. On the reverse size series she was able to switch over more quickly than with colour although she did not show immediate polar cross-over (A choice to D intervened). It is possible that she chose

¹That the “reverse” colour series, the subject’s fourth series, required the most trials to learn is interesting with respect to recent findings on multiple list learning by rhesus monkeys (Swartz, Chen, & Terrace 1991). They found that for rhesus monkeys the fourth list seemed more difficult than preceding lists, however they were unable to discern why this was the case.

this item based on the fact that in Experiment 3.1 the absolute value of item D, whilst second biggest in this series, served as the first item (biggest) of the "test set" in over-training. After this failed, she made the correct polar cross-over choice to A (smallest). However, it is also possible that the choice to D was part of an exhaustive search and that her subsequent choice to A was by chance. Secondly, in the trial blocks tests, subjects had little difficulty in switching back to their "original" monotonic size or arbitrary colour series. Errors were typical one-step forward errors in both cases. Finally patterns of inter-item RT indicate that subjects used phrasing to report both monotonic size or arbitrary colour series. This use of phrasing for both types of series may indicate that subjects treated both as arbitrary.

Summary of Experiments 3 & 4

Intensive training on a five-item touch screen analog of the seriation task did not result in success in most subjects. Although there was some learning, task modifications designed to better communicate the task to the subject, such as parsed input and immediate feedback, did not result in young subjects' being able to reach the success criterion with as few as five items. Most subjects were limited to the control of at most four items, not only for monotonic size series, but also for an arbitrary colour string. For young subjects there was no selective performance in favour of the monotonic size series that would indicate that subjects were using dimensional control to manage the series. Both arbitrary and non-arbitrary series appeared to be have been controlled in the same way. These limitations pointed to an unexpected constraint on serial control in three year olds. From this it could be concluded that for young subjects, failure at seriation is at least the result of limitations at serial control *per se*. Error patterns also indicated that subjects were using low-level exclusion strategies to manage short series.

There were three subjects however who were able to reach the full five-item criterion for both series. Through examination of concurrent performance on both types of string and examination of performance on transfer to the reverses of both arbitrary and non arbitrary strings, we were able to disambiguate this success. We found that subjects were not likely to be using dimensional control to report the series. The monotonic size series was being reported through the use of serial control in much the

same way as the arbitrary colour series. Analysis errors and reaction time indicated that subjects may also have been using phrasing to optimize search through both series.

Although at the outset of the touch screen based assay we knew very little about the robustly failing age group of subjects, by the end we knew considerably more about why these subjects fail at seriation and what their performance limitations were. The implications of these findings, coupled with those from the extended clinical assay, for a characterisation of seriation and of developmental in general are discussed in the next chapter.

Chapter 7: Conclusions and issues for further research

This thesis set out to contribute to an eventual characterisation of causal factors in development by providing a rich description of developmental phenomena. In this case the task domain was seriation. The general goal was to make a start at understanding how and why young children fail at seriation, not merely reporting that they do. This problem was tackled in two experimental phases, the extended clinical assay and the touch screen based assay. In the extended clinical assay the main findings were: (1) intervention, for the most part, did not change performance, (2) subjects under age four years were entirely resistant to success, and (3) some older subjects were able to succeed with assistance. In the touch screen based assay, subjects under age four years were focussed on for intensive training. It was found that subjects continued to be resistant to success and that the source of their limitations was *at least* a limitation of serial control.

In this chapter conclusions and implications arising from these findings are discussed. However, before discussing the merits of the thesis it must be said that there were a number of limitations on these studies which, of course, limit their scope in terms of conclusions which can be drawn from the data and implications for a general understanding of development.

Firstly, in this thesis a micro-analytic approach was taken. Small numbers of subjects were used and the primary focus of research was restricted to a narrowly banded age range of young subjects. The aim in keeping to smaller numbers was to be able to deeply probe a particular group to discover, as fully as possible, the extent of their abilities with respect to the decomposition of seriation. In particular we sought to locate a target group where real progress in terms of information gain could be made. Of course this sacrificed the generality offered by a large scale macro-analysis. However the motive was that the depth value of the information gained would justify this trade-off. Secondly, having located the target group, certain further restrictions with respect to the extent to which the

decomposition skills could be probed had to be made as well. We focussed on serial production aspects of seriation, in particular, the control of arbitrary and non-arbitrary strings. Not having assayed the full range of skills in the decomposition, we are limited in what we can say about the totality of young subjects' abilities and limitations. However serial production was chosen because of its likely role as a basic skill and for the possibility of comparison with existing literature. For these reasons it had great promise as an inroad to discovering an important part of the baseline of seriation ability in young subjects.

In this chapter, the benefits of this approach are discussed, firstly in terms of conclusions about the development of seriation and secondly in terms of implications for the characterisation of developmental change as a whole. Finally we address the prescriptive value these conclusions have to offer further research.

Conclusions about seriation

Given the investment in the extensive micro-analytic experimental programme in this thesis, how does it get beyond existing descriptions of seriation?

Genuine competence limitations

This thesis contributes to a consensus about the reliability and timing of age-related stages in development. The general time scale of the stages and changes in behaviour on seriation tasks did not crumble under experimenter intervention. This is not to say that intervention and training resulted no change whatsoever. There were age-related differences in the uptake of feedback. Some older subjects were able to change through the kinds of social intervention provided in the extended clinical investigation (Experiment 1). The implications of this change is discussed below. Young subjects (under age four years) exhibit some change with protracted instruction, but the scope of their change is very limited.

Recall that for these young subjects instruction was of two types. In Experiment 1 they received social types of intervention. This intervention was in the form of verbal interactions with the experimenter. Subjects

were given the opportunity to review their productions and compare them with the model. Hints of increasing strength regarding the location and repair of errors were offered to the subject. In Experiment 3 they received traditional "learning paradigm" types of instruction. Training was conducted non-verbally as far as was possible. The input was parsed so that subjects seriated smaller numbers of objects first. The set size was incremented only after a subject had reached a training criterion. Error was signalled immediately by an unambiguous "bleep."

Neither of these instructional methods lead young subjects to lasting success. In Experiment 1, the under fours were not helped by opportunities to review their productions and hints provided by the experimenter. They were not self-motivated to effectively repair their productions and neither could they be socially motivated. In Experiment 3, the majority of subjects were unable to reliably control as few as five items, despite training using parsed stimuli and immediate feedback.

These results taken together have several important implications. The seriation failures of young subjects are robust and real. There appear to be genuine competence limitations in subjects younger than four years, based on elementary seriation tasks. Experience is not a panacea. Despite attempts to facilitate successful solutions by intervening socially and despite intensive training using tried and true learning methods, young subjects continued to fail. These results therefore do not support the view that young children's failures are of a performance nature and could be got round by "negotiating meaning" with the child. There seem to be real changes in competence occurring over the time course of the development of seriation.

These results confirm that the results presented by Inhelder and Piaget (1964) and others are reliable and thus establish seriation as a reliable index of change. This is important if seriation is to be further explored. Indexicality has been a problem for other task domains, such as linear transitive inference and conservation (see reviews by Breslow, 1981 and McShane, 1991), which seem at first blush to be good candidate domains for the study of cognitive change. Given the position that there are real changes in competence in development, if seriation were not a reliable index of change then it would be of little use as a "proxy" for development

in general. Thus further extended analysis of seriation behaviour designed to make transparent the causal mechanisms of change are justified. Also, further results can be situated within a context of long-term cognitive change.

Possible proximal zone

The reader will recall that social intervention provided in Experiment 1 did result in change in behaviour for some older subjects. As per Vygotsky's (1978) notion of a "zone of proximal development," there appear to be phases in development, before children can independently and spontaneously succeed at a task, where they are able to take advantage assistance provided by an adult to achieve a successful solution. Whilst some of these "assisted successes" were short-lived, (that is, when assistance was withdrawn the subject reverted to failure), other subjects achieved long-lasting improvements in their performance. We called these subjects "transitional successes." The highest concentration of transitional subjects was found in the six months between ages five-and-a-half and six years. It is interesting to note that this is toward the end of a left hemisphere growth spurt (between ages four and six years) outlined by Thatcher et al., (1987) (more will be said about this later). That some subjects were able to succeed after intervention implies that instruction can play an important role in cognitive change at certain times in development. There appear to be certain age "zones" when subjects are ripe for instruction. However, that there are subjects who cannot improve or who improve only temporarily indicates that instruction is not the only factor underwriting change. Experience appears to be a necessary but not a sufficient condition for change.

Resistance points

The purpose behind providing intervention and intensive training was not only to see whether children could improve their performances on seriation tasks. Intervention also provided an opportunity to diagnose resistance points to success. Having demonstrated that young subjects show genuine competence limitations, the next step was to try and specify what those limitations were.

In the extended clinical investigation the role error information plays at different points in development was examined. Of particular interest were young children, especially those under four years of age, who were completely resistant to success. Although it is likely that there were many factors contributing to their failure, of the error factors focussed on here, the main resistance point in young children was found to be a lack of awareness of error. Young children under age four years did not appear to register error, neither by themselves by direct comparison of their own productions with the model, nor did hints from the experimenter have a lasting effect.

How is it that these young subjects were not aware of their default? Lack of error awareness may have implications with respect to the child's own characterisation of the seriation task. Young children may not characterise the seriation task in the same way that older children or adults do. It is possible that children saw their own inadequate productions as acceptable solutions. As Sonoda points out, young subjects may see the goal of a seriation exercise as just arranging the sticks in a line. In this case it is not surprising that young children were resistant to success, as they had a different image of achievement. Neither could they be prompted into success when provided with a social motivation to change --- the signalling of mistake and hints provided by the experimenter. Finally, even when provided with the unambiguous signalling of error in the touch screen analog of the seriation task, subjects showed some improvement in seriation but were limited in their item capacity to four items at best.

Although it makes a good start in descriptive terms, it is not enough to merely say that subjects fail because they don't have a faulty image of achievement or an inadequate characterisation of the task. It is necessary to get beyond this to discover why young subjects have inadequate task characterisations and why they behave in the way that they do.

Recall that Experiment 4 focussed on making more transparent the failure of these young subjects. The touch screen and associated paradigms were used to explore possible sources of error. Here the concentration was on serial control (as per the McGonigle and Chalmers decomposition). An

unexpected age constraint on serial control was found. This failure of serial control among three year old subjects was unanticipated. It had been hypothesized that serial control would be intact in these young children, because serial learning has been reported in non-human species such as the pigeon (Straub & Terrace, 1981) and monkey (D'Amato & Colombo, 1988), and results reported in Chapter 4 showed that four year olds can learn a five-item colour series using two training conditions. Failure in subjects only six months to a year younger was surprising. When results from both Chapters 4 and 5 are combined, a "cutoff" for serial learning of a five-item set can be seen at about three-and-a-half years of age. It is possible that crucial changes related to serial ordering ability are happening within a small metric of time, perhaps within only a few months. Serial ordering tasks may be an important age-sensitive index of cognitive change in young children.

From these results it can be concluded that performance limitations on seriation in these young subjects are at least limitations of serial control. Considering that these young subjects had trouble controlling as few as four familiar and discriminable items (coloured squares) in a series, it stands to reason that they should fail to control ten blocks differing only slightly in size. Of course there may be other factors contributing to this failure. With respect to the McGonigle and Chalmers decomposition, the ordinal computation abilities of these young subjects may be called into question. Subjects had difficulty reporting two five item size sets in Experiment 3.1. This may indicate that their ordinal computation abilities are not intact. The subjects tested by McGonigle and Chalmers (1986) were slightly older, and as we have seen, a few months can make a difference. This is a matter for further investigation.

Returning to young subjects' characterisation of the seriation task, that young children show similar performances (in terms of item limitation) on both ordinally related and unrelated series suggests that they may have been looking at them as the same kind of problem. Young three year olds who were limited to the control of only three or four items very likely to see both size and colour series problems in terms of exclusion principles. Older subjects between three-and-a-half and four-and-a-half years old were likely to have seen both problems as serial control problems only, treating the elements of the size series as arbitrary elements unrelated to each

other. The lack of these skills seems to have had an influence on how subjects characterised and found solutions to the tasks before them. This is quite different from the Piagetian notion that the seriation task is obvious to the child (owing to its "concrete" nature). By his characterisation developmental changes involved children recruiting new means of solving, what is to them, the same problem. However it is likely that because children lack certain means, they may see a different task. Thus, any attempt at characterising changes in development must take both factors into account --- children's means of solving a task in conjunction with their developing characterisation of it.

Finally, findings here indicate that important changes in serial and seriation skills occur over a time scale of perhaps only a few months. Recall that for serial ordering, changes appear to be happening around age three-and-a-half years. For seriation, as seen in the extended clinical assay, important changes occur around age five-and-a-half years. This would emphasize the importance of looking at individual subjects micro-analytically, if crucial changes are to be captured and characterised.

Comparison to the Piagetian account of seriation

The path of development for seriation outlined here is one characterised by phases of great change, a refutation of prior strategies and solutions, followed by phases of repair, honing and perfecting new strategies and solutions. How does this compare with the Piagetian account (1952; Inhelder & Piaget, 1964)? Although there is agreement on the basic time scale for the development of seriation, there are many differences between the Piagetian account and the one presented here.

Piaget didn't investigate whether or not seriation could be taught. He focussed on self-discovery and favoured scenarios which maximized this possibility for children. However results presented here indicate that children may achieve success with assistance, just before Piaget's stage of trial and error success. Thus some transitions may come about through external prodding rather than self-discovery alone.

Further, Piaget's characterisation of seriation rests on the acquisition of the key rule of reversibility, with the endstate of seriation being the ability to use this operation to coordinate a series of size relations. The progression

of seriation behaviour is always explained in terms of the coordination of relations. As a consequence, Piaget's stages of seriation development form a continuous line. Early inadequate solutions are presented as approximations of the endstate. Young (1976, 1978) presented a similar view. He concluded that early inadequate seriation attempts did not arise from distinct pre-seriation processes but were the result of the subject's lack of one or more critical rules.

The data presented here indicate that the reasons for default are multifactorial. Further, these reasons may be different at different points in development. As new competences enter the system, early strategies and solutions are refuted in favour of new ones. Early inadequate solutions are not necessarily related in a causal way to later successful solutions.

Take for example Piaget's stage of uncoordinated series. In Piaget's terms uncoordinated series results from isolated size comparisons which the subject cannot coordinate into one series. The subject's production is therefore two or three small series displayed in a line. However this kind of series could be the result of a faulty task characterisation (as discussed above). Random selection and placement on items might result in the appearance of "small uncoordinated series," without any attempt by the subject to incorporate "size" into the reproduction of the model. In this way it is possible that the production of "small uncoordinated series" has nothing to do with the coordination of size relations. Further, it is difficult to see how this kind of solution would "scale up" to include attention to size relations. Later when subjects are achieving what Piaget called "trial and error success," it is likely that they have discarded their old characterisations and solutions for something entirely different. By this time serial ordering and ordinal computation skills are likely to be intact, although not necessarily working in concert (McGonigle 1987). With these new competences at their disposal subjects have a new basis on which to build their strategies and solutions. In this way we see that whilst styles of performance in younger subjects, of course, indicate the solution of the day, they are not necessarily proto-versions of the final endstate solution.

Conclusions about developmental change

The reader will recall discussion of recent independent converging evidence about discontinuities in development provided by Thatcher et al. (1987), who have shown neurophysiological evidence of discontinuities and growth spurts in cerebral hemispheric development during childhood and adolescence. The temporal boundaries of the hemispheric growth spurts cross-correlate with the temporal boundaries of stage-like changes in cognitive behaviour reported by Piaget and the Genevan group. In particular there is a left hemisphere growth spurt between ages four and six years. These results and characterisations combine with those presented in this thesis to suggest a developmental scenario which is essentially a two factor process. Firstly, there are fundamental changes in competence which are "engineered" by genetic pre-programming. These occur at pre-specified points along the developmental time scale from birth to adolescence. Secondly, through the child's discovery and experience, new competences are implemented. These implementations are expressed behaviourally as qualitatively different performance at different periods (stages) in development. This approach suggested here is not just ontogenetic or behavioural but synthetic, taking into account both biological and behavioural factors in development. Thus development may reasonably be seen as a progression of qualitative changes in competence which may be discontinuous with previous levels of competence.

This approach has many implications. This approach is quite different from theories which call for increasing access to a fixed competence. Here it is not the "access" which increases, but the competence itself. An important implication is that there is no obligation to trace a continuous line from early performance to later performance on a particular task or skill.

Similarly for stage accounts, the experimenter is relieved of the obligation to explain how the behaviour of one stage yields up the next. Because new competences may be entering the system, early behaviour and solutions may be based on different "hardware" than later behaviour and solutions. Early behaviour need not be causally linked to and continuous with later behaviour.

With respect to the role of instruction and training, an implication of the two factor scenario is that without the relevant underlying competence in place, training (however intensive) has only limited effect. Only when the relevant competence is already in place is instruction a sufficient condition for change. Thus, the necessary and sufficient conditions for change are that the relevant competence(s) are in place and that there is the opportunity for their implementation. Several implications arise. If change occurs slowly over a long period of time it could be because the underlying competence is missing. In this case, one would expect to find resistance points to success in the case of instruction. If changes occur within a short period of time, then there are two possibilities. Either both conditions have been met or, given instruction, rapid changes in behaviour may be the result of mimicry. In first case, one would expect behaviour changes to be permanent. In the second case, the underlying competence may be missing. Behaviour changes would be expected to be non-robust. Further, there may be a time lag between the "delivery" of new competences and their behavioural expression. During this time subjects may be able to succeed with assistance, but perhaps not spontaneously.

Using the left hemisphere growth spurt between ages four and six years (Thatcher et al., 1987) as a guideline, the data are quickly reviewed in light of the above implications. All subjects under age four, before the growth period, changed only slightly and were resistant to success. Resistance points included inadequate task characterisation, lack of error awareness, and limitation in serial control. Many subjects between four and five and a half showed non-robust changes (the short-lived category in the extended clinical assay). Finally, recall that "transitional subjects" were found between age five-and-a-half and six years, toward the end of the period of growth.

Thus, our results are consistent with the two factor approach in which the necessary and sufficient conditions for human cognitive development are the genetically pre-programmed unfolding of specific brain functions and the opportunity for their implementation once a new bit of neural competence has been delivered. These findings outline a multi-faceted and non-linear path of development, which is a more richly described a

potentially useful advance upon what we already knew, that is, that development is characterised by age-related stages of change.

Issues for further research

The shift from blocks to the touch screen

In this thesis, a move was made away from traditional materials (blocks) used for studying seriation to touch screen technology. The touch screen was found to have several advantages over blocks which would recommend it for further research on seriation. Firstly, the touch screen has advantages with respect to maintaining subjects' interest. The children very much liked working with the touch screen and maintained an interest in the tasks. This was important considering that intensive training meant daily sessions over the course of months. It was important to have a task which held the children's interest in order that they would keep coming back for their daily sessions and would dedicate their attention to the task at hand. That it was possible to maintain the interest of children as young as three years old was an important achievement.

Secondly, the touch screen allows for non-verbal administration of tasks. Because feedback is communicated by unambiguous "bleeps," there is little question of whether or not children understand what is meant. Also because verbal instructions from the experimenter can be kept to a minimum, there is less chance of the subject misunderstanding the instructions. In this respect the touch screen can be used with very young children who do not yet have a command of language. The reader will recall that the touch screen has been used profitably with non-human subjects. (e.g. Terrace, 1987).

Thirdly, the touch screen has advantages with respect to control of the experimental situation. Because stimuli can be quickly and automatically generated, there were no delays for the experimenter having to set up a new array of blocks. Such delays can cause the subject's attention to wander. Further, there are no opportunities for the subject to disturb an array, whereas blocks can be mischievously moved. Rapid presentation of trial arrays allows for the best use of the subject's time. The data presented in this thesis were based on hundred of trials collected per subject. Collecting the same amount of information with blocks would have taken

a great deal more time, and it is likely that it would not have been possible at all. Further, because the experimental situation was controlled by a computer, it was possible to arrange for the automatic collection of data and also to get accurate measures of reaction time which would have been very difficult in the block scenario.

Finally, the touch screen is versatile. It was possible to use the touch screen to test many different aspects of seriation without a great deal of difficulty. For example, the number of items, as well as their sizes, can be easily manipulated. Further, it is possible to manipulate the nature of type of the items. Here different colours and sizes were used, however it is possible to use sets of different shapes or even pictures.

For these reasons, it seems that using blocks for studying seriation is not the best course for future research. Although, it is likely that blocks are still the best means of looking at certain surface aspects of seriation, such as the subject's manipulation of items in "trial and error" phases of seriation. However, for reason discussed above, such studies are unlikely to get at the causal mechanisms behind the development of seriation. In order to get a look at the component skills which make up seriation and the interplay between them, the touch screen appears to be the way forward.

Further exploration of young subjects' limitations

Above it was concluded that young subjects' limitations were at least limitations of serial control. However, it is possible that other important skill defaults were contributory to their failure. The reader will recall modelling problems found in the extended clinical assay. Of particular interest is subjects' ordinal computation skills. Was the ordinal computation component intact in these subjects? It is likely that it is not, due to the fact that subjects given two different size sets had difficulty switching between them. Nevertheless it is important to establish which skills are present and which are absent in young subjects in order to have a clear baseline upon which to base findings with older subjects.

Recall the three subjects in Experiments 3.3 and 3.4 who seemed to be equally competent at serial order tasks with arbitrary colour stimuli and monotonically sized stimuli. Although it was reasonably clear that these

subjects were not using dimensional control to report the size series, certain questions remained. Serial order seemed to be intact in these three subjects, but what about their ordinal computation skills? It may have been the case that both serial ordering and ordinal computation skills were intact in these subjects, but that the lack of interaction between these two skills was the source of their lack of dimensional control (see McGonigle 1987). On the other hand it may be the case that ordinal computation is not intact and that this is the source of their default. Further exploration of subjects aged three-and-a-half years, a possible lower bound for serial ordering and ordinal computation skills, is required.

Also, it must be remembered that in this experiment there was a limitation that no more than five items could be presented to the subject. It may be the case that a series of five items is an insufficient number to "bring out" differences in the reporting of arbitrary and non-arbitrary series. With small numbers of items subjects may "revert" to simple strategies, such as rote memorization. Indeed, Piaget chose ten items as the set size for classical seriation. It may be that he knew that such a number was required in order for relational skills to manifest themselves. In order to better test serial ordering and the possibility of dimensional control, subjects must be pushed to the ends of their item capacity.

Exploration of subjects on the brink of success

Most of this thesis concerns the dimensions of failure in very young children. Having made a start at a rich, micro-analytic description of the early development of seriation, the next step is to investigate subjects who are nearer the mark with respect to seriation success.

A specific point of interest is the development of dimensional control. Subjects who are able to take advantage of dimensional information to search through a monotonic string have an immense advantage in terms of reducing memory load and can therefore have a very large item capacity. Without this appreciation, as we have seen, subjects have a very high memory load which limits item capacity. Such subjects act upon the monotonic series as they would an unrelated set of stimuli such as the colour string. If subjects are further limited in serial control, they resort to probabilistic exclusion strategies to report the string. Here the subject

becomes more sharply limited in terms of set size. However, in the course of development one would expect that eventually the monotonic size string would begin to have an advantage over the arbitrary string. Piaget did not sufficiently point out the specialness of the monotonic size string. His "coordination of relations" solution did not take into account that adjacent size relations could be non-monotonic.

In the extended clinical phase some preliminary work was done using monotonic and non-monotonic sets with successful seriators (see Appendix C). Subjects were asked to copy both monotonic and non-monotonic models. Subjects access to review models was limited. It was found that most subjects could reproduce the monotonic model with one short viewing, whilst non-monotonic models required repeated review (and even after repeated review they were not always able to reproduce the non-monotonic series). Subjects were further required to model series which were only minimally degraded from the monotonic (two adjacent blocks swapped). Here it was found that subjects typically erred in favour of the monotonic. Thus once the subject develops an appreciation for dimensional information they can use it to reduce memory load.

Further manipulation of both set size and string quality are needed if we are to discover which strings facilitate sophisticated means of control and how these means of control emerge. It would be particularly interesting and beneficial to pinpoint subjects who are just below the transitional zone (between five and five-and-a-half year old) for a study of this type.

Closing

The findings presented here outline a multi-faceted and non-linear path of development, which is a more richly described and potentially useful advance upon what we already knew, that is, that development is characterised by aged-related stages of change. Having moved toward a descriptive base which will support a causal account of transitions in development, the question of how a weak system becomes a strong system remains. Findings in children's brain growth by Thatcher et al. (1987) found that bursts in brain growth are correlated with the timing of Piaget's stages. This suggests that there may be a maturational component driving change. Indeed, whilst behavioural changes are taking place the system itself is growing and changing. In this way development can be seen as

being driven partly by maturational and partly by behavioural factors. Bursts in brain growth inject more power into the system, the system then takes advantage of this boost, refuting old strategies and beginning to develop new ones. This ushers in a phase characterised by repair in which new strategies are honed via the familiar mechanisms of learning. In this way, building up a multi-faceted picture of development, we begin to tease apart the factors behind why we find stages in development.

Although there is still a great deal of work to be done, this work is a step toward understanding transitions in development and thereby understanding how a cognitive system boots itself up. How systems "invest in complexity" and become more powerful is a major epistemological issue spanning formal systems research, artificial intelligence, robotics, and biological systems research. This research opens another window into this pervasive and all-important problem.

References

- Breslow, L. (1981). Reevaluation of the Literature on the Development of Transitive Inferences. *Psychological Bulletin*, 89, 2, 325-351.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press
- D'Amato M. R., & Colombo, M. (1988). Representation of serial order in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 131-139.
- D'Amato M. R., & Colombo, M. (1989). Serial learning with wild card items by monkeys (*Cebus apella*): Implications from knowledge of ordinal position. *Journal of Comparative Psychology*, 103, 252-261.
- D'Amato M. R., & Colombo, M. (1990). The symbolic distance effect in monkeys (*Cebus apella*). *Animal Learning & Behavior*, 18, 133-140.
- Dodwell, P. C. (1960). Children's understanding of number and related concepts. *Canadian Journal of Psychology*, 14, 191-205.
- Dodwell, P. C. (1961). Children's understanding of number concepts: Characteristics of an individual and of a group test. *Canadian Journal of Psychology*, 15, 29-36.
- Donaldson, M. (1978). *Children's minds*. London: Fontana.
- Donaldson, M., Grieve, R., & Pratt C. (1983). *Early childhood development and education*. Oxford: Basil Blackwell.
- Elkind, D. (1961a). The development of quantitative thinking: A systematic replication of Piaget's studies. *The Journal of Genetic Psychology*, 98, 37-46.
- Elkind, D. (1961b). Children's discovery of the conservation of mass, weight, and volume: Piaget replication study II. *The Journal of Genetic Psychology*, 98, 219-227.
- Elkind, D. (1964). Discrimination, seriation & numeration of size & dimensional differences in young children: Piagetian replication study VI. *The Journal of Genetic Psychology*, 104, 275-296.
- Emde, R. N., & Harmon, R. J. (1984). Entering a new era in the search for developmental continuities. In R. N. Emde & R. J. Harmon (Eds.), *Continuities and discontinuities in development*. New York: Plenum Press.

- Flavell, J. H. (1963). *The developmental psychology of Jean Piaget*. Princeton, NJ: Van Nostrand.
- Gelman, R. (1978). Cognitive development. *Annual Review of Psychology*, 29, 297-332.
- Gelman, R. & Gallistel, C. R. (1978). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.
- Inhelder, B., & Piaget, J. (1964). *The Early Growth of Logic in the Child*. London: Routledge & Kegan Paul.
- Kagan, J. (1984). Continuity and change in the opening years of life. In R. N. Emde & R. J. Harmon (Eds.), *Continuities and discontinuities in development*. New York: Plenum Press.
- Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition*, 23, 95-147.
- Lashley, K. S. (1960). The problem of serial order in behaviour. In F. A. Beach, D. O. Hebb, C. T. Morgan, & H. W. Nissen (eds.), *The Neuropsychology of Lashley*. Ch. 30, New York: McGraw-Hill, pp. 506-528.
- Light, P. (1986). Context, conservation, and conversation. In M. Richards & P. Light (Eds.), *Children of social worlds*. Cambridge: Polity.
- Light, P., Buckingham, N., & Robbins, A. H. (1979). The conservation task as an interactional setting. *British Journal of Educational Psychology*, 49, 304-310.
- Light, P., & Gilmore, A. (1983). Conservation or conversation? Contextual facilitation of inappropriate conservation judgements. *Journal of Experimental Child Psychology*, 36, 356-363.
- Marshall, J. C. & Morton, J. (1978). On the mechanics of EMMA. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language*. Berlin: Springer-Verlag.
- McGarrigle, J., & Donaldson, M. (1975). Conservation accidents. *Cognition*, 3, 341-350.
- McGonigle, B. O. (1987). Nonverbal thinking by animals? *Nature*, 325, 110-112.
- McGonigle, B. O., & Chalmers, M. A. (1986). Representation and strategies during inference. In T. Myers, K. Brown, & B. O. McGonigle (Eds.), *Reasoning and Discourse Processes*. London: Academic Press.

- McGonigle, B. O., & Chalmers, M. A. (in press). *Intelligent systems: A cognitive analysis*. Columbia University press.
- McShane, J. (1991). *Cognitive development: An information processing approach*. Oxford: Basil Blackwell.
- Newell, A. & Simon, H. (1972). *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- Piaget, J. (1970). Extracts from Piaget's theory. In P. H. Mussen (Ed.), *Manual of child psychology*. London: Wiley.
- Piaget, J. (1974). *The child and reality*. London: Frederick Muller Ltd.
- Piaget, J. (1952). *The Child's Conception of Number*. London: Routledge & Kegan Paul.
- Sonoda, N. The development of self-corrective activity in serial ordering. Unpublished manuscript, Kyushu University.
- Straub, R. O., Seidenberg, M. S., Bever, T. G., & Terrace, H. S. (1979). Serial learning in the pigeon. *Journal of the Experimental Analysis of Behavior*, 32, 137-148.
- Straub, R. O., & Terrace, H. S. (1981). Generalization of serial learning in the pigeon. *Animal Learning & Behaviour*, 9, 454-468.
- Sugarman, S. (1987). The priority of description in developmental psychology. *International Journal of Behavioral Development*, 10, 391-414.
- Swartz, K. B., Chen, S., & Terrace, H. S. (1991). Serial learning by rhesus monkeys: I. Acquisition and retention of multiple four-item lists. *Journal of Experimental Psychology: Animal Behavior Processes*.
- Terrace, H. S. (1987). Chunking by a pigeon in a serial learning task. *Nature*, 325, 149-151.
- Terrace, H. S. (1991). Chunking during serial learning by a pigeon: I. Basic evidence. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 81-93.
- Thatcher, R. W., Walker, R. A., & Guidice, S. (1987). Human cerebral hemispheres develop at different rates & ages. *Science*, 236, 1110-1113.
- Vygotsky, L. S. (1934). *Thought and Language*. (Trans. by E. Hanfman & G. Vakar). Cambridge, MA: MIT Press, 1962.

- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Ed. M. Cole, V. John-Steiner, S. Scribner, & E. Souberman. Cambridge, MA: Harvard University Press.
- Wertsch, J. V. (1985a). *Vygotsky and the Social Formation of Mind*. Cambridge, MA: Harvard University Press.
- Wertsch, J. V. (Ed.). (1985b). *Culture, communication, and cognition: Vygotskian perspectives*. Cambridge: Cambridge University Press.
- Young, R. M. (1976). *Seriation by Children: An Artificial Intelligence Analysis of a Piagetian Task*. Basel: Birkhauser.
- Young, R. M. (1978). Strategies and the Structure of a Cognitive Skill. In G. Underwood (ed.), *Strategies of Information Processing*. London: Academic Press, 357-401.

Appendix A: State Evaluation Study Procedures and Results

EXPERIMENT

Subjects

8 three year olds at the Edinburgh University Psychology Department Nursery.

Materials

2 sets of 4 cards each printed with photographs of series of 5 cubes: monotonic ascending, monotonic descending, maximally degraded non-monotonic (4, 3, 5, 1, 2, where 1 is the largest block and 5 is the smallest), minimally degraded non-monotonic (1, 3, 2, 4, 5). Also, the set of 5 blocks from which the photographs were taken and an identical additional set.

Design

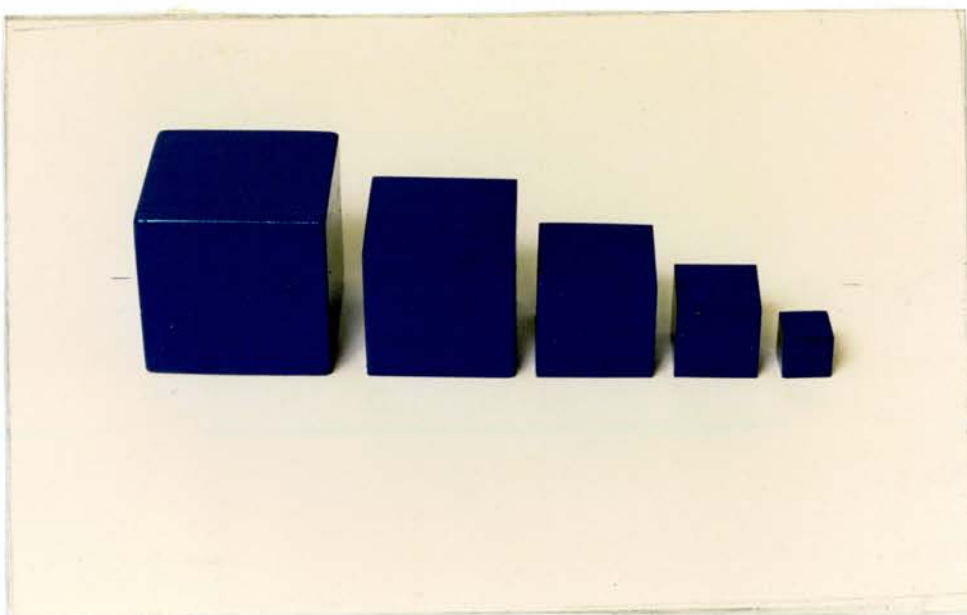
The experiment was divided into two parts. In the first part, the subject was required to verify whether a photograph of a configuration of blocks was the same as its corresponding three-dimensional construction. In the second part, the subject was required to register disparity between two photographs, globally in the first instance, and latterly in more specific ways.

Procedures

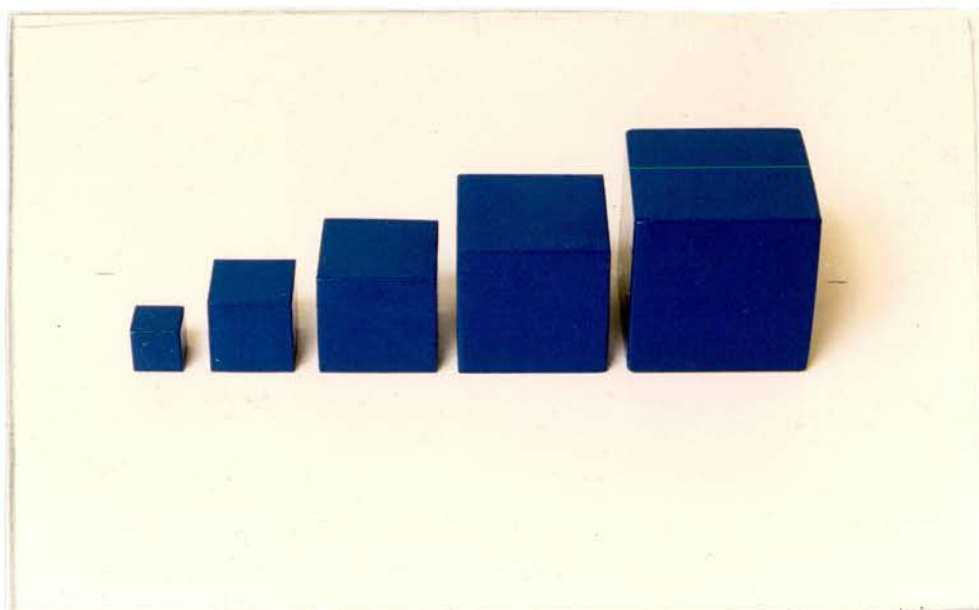
Verification

Stimuli: 1 set of 4 cards each printed with photographs of series of 5 blue cubes. 1 set of 5 blocks (from which the photographs were taken).

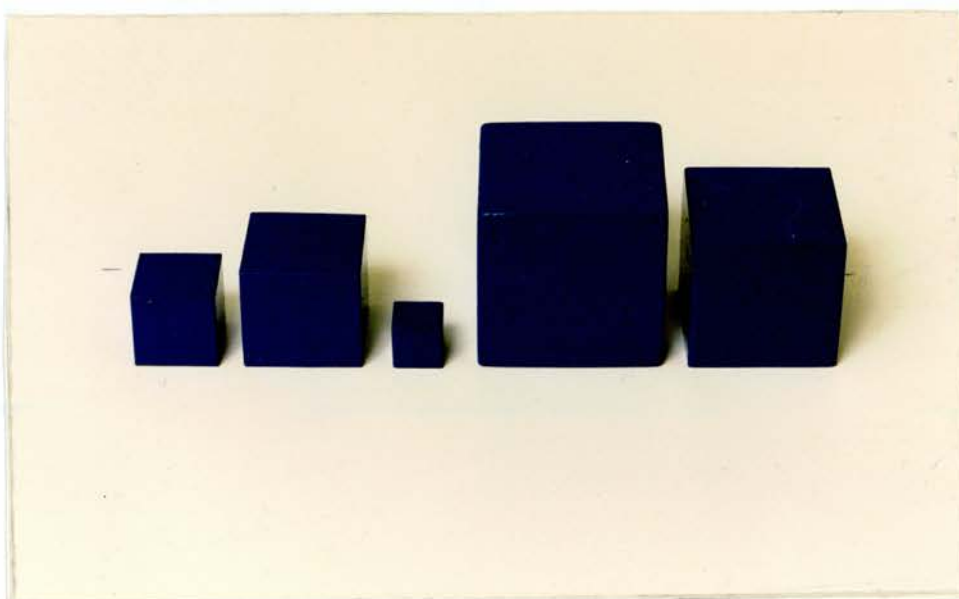
Treatment: Subjects were required to verify whether a pictured series of blocks matched a actual block series on the table. Successive presentations of the four photographs were made once each in a quasi-random order. Subjects received four trials, i.e. one exposure to each of the four block configurations. Trials were presented in a quasi-random order.



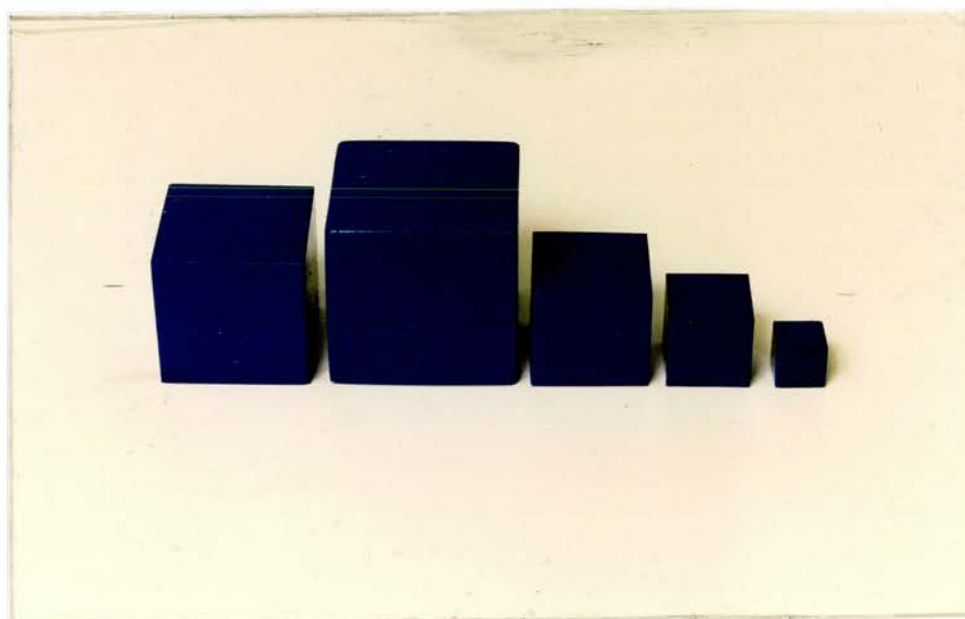
The Monotonic Descending Series (DSC)



The Monotonic Ascending Series (ASC)



The Non-Monotonic Series (MAX)



The Degraded Series (MIN)

The subject entered the testing room where there was a set of five blocks spread out on the table (the set five blocks were those used in the photographs). The experimenter introduced the session as follows. "Look at these blocks here. You can make things with blocks, etc." The experimenter then removes the photographs from view and introduced the task as follows. "I've got these blocks here and some pictures of blocks too. I'm going to show you some blocks and then show you a picture of some blocks and you tell me if the blocks in the picture are just like the blocks here." Behind a screen, the experimenter then constructs (one of the four possible) series. The use of a screen prevents any serial production notions from being introduced to the subject surreptitiously. The subject was then presented with a picture (under an acetate sheet) and asked "Is the picture just like the blocks here?" Successive presentations of the four possibilities were made once each in a quasi-random order. The subject's choice was logged. After the subject has had the opportunity to respond to each of the four pictures, the array was removed and the trial was completed. The procedure was then repeated with the next block series. All sessions were video taped. Data were also recorded on protocol sheets ticked off in situ.

Disparity Recognition and Identification

Stimuli: 2 sets of 4 cards each printed with photographs of series of 5 blue cubes.

Treatment: Subjects were required to indicate whether two pictures were the same or not. One picture was deemed to be "the subject's" and the other was deemed to belong to a confederate teddy bear. Subjects were required to register disparity in two directions to assess the subject's ability to reverse topic and comment. In cases where disparity was registered, subjects were required to point out the specific location of the disparity. There were ten trials (see table below) which were administered in a quasi-random order.

ASC-ASC	ASC-DSC	ASC-MAX	ASC-MIN
# # #	DSC-DSC	DSC-MAX	DSC-MIN
# # #	# # #	MAX-MAX	MIN-MIN
# # #	# # #	# # #	MAX-MIN

"Disparity recognition" task trials

The subject was seated at a table with the experimenter seated across from him or her. Also on the table was a large teddy bear. The experimenter explained that, "Teddy has some pictures and so do you. Can you tell me if your pictures are just like Teddy's." Two pictures were presented simultaneously (under an acetate sheet). First the experimenter asked the subject if his or her picture was "just like" Teddy's. The subject answered and the experimenter noted the truth or falsity of the response. The subject was then asked, "Is Teddy's just like yours?" If the subject's answers to both these questions agreed, then the experimenter presented the next trial. If answers to both these questions did not agree, then the subject was asked again, "Is yours just like Teddy's?" The response was noted and the experimenter presented the next trial. In cases where there was agreement that the two pictures were different (on both questions, whether it was indeed correct or not) the subject was required to indicate "where are they different" before going on to the next trial. All sessions were video taped and data were recorded on protocol sheets ticked off in situ.

Results

Verification

MODEL

MATCH

	ASC	DSC	MAX	MIN
ASC	100%	13%	50%	50%
DSC	25%	100%	50%	38%
MAX	63%	50%	100%	50%
MIN	63%	63%	50%	100%

Percent correct (mean for 8 subjects)

All subjects could correctly verify all four block configurations with their corresponding pictures when asked whether they were the same. However, three of our youngest subjects responded "yes" to every verification query, i.e. that each block configuration was the same as each of the four pictures. Another subject similarly responded "yes" to all verification queries when the models were non-monotonic. Among the remaining subjects, the ability to verify that block configurations were the same as their corresponding pictures was good with the exception of three

particular comparisons. Subjects typically confused the monotonic ascending with the monotonic descending. Also subjects typically confused the monotonic descending with its minimal degradation, but more so when the minimal degradation served as the model.

Disparity Recognition

TEDDYS

<i>YOURS</i>		ASC	DSC	MAX	MIN
	ASC	100%	0%	50%	38%
	DSC	XXX	100%	50%	38%
	MAX	XXX	XXX	100%	38%
	MIN	XXX	XXX	XXX	100%

Percent correct (mean for 8 subjects)

All subjects responded correctly to queries involving identical pictures. However, the same four subjects as above answered "yes" to every query. Two of these were the only subjects to have difficulty with topic-comment reversal. Among the remaining subjects, ability to recognise disparity was good with the exception of four particular comparisons. No subject differentiated the monotonic ascending from the monotonic descending. Also poor performance was found in comparisons involving the minimally degraded configuration. All subjects, except one, indicated specific location of disparity between block configurations when queried. All specific locations were given in terms of pole blocks, i.e. the biggest or the smallest block.

Appendix B: Seriation with Reduced Sets

EXPERIMENT

Subjects

14 four year old children at the Edinburgh University Psychology Department Nursery.

Materials

2 sets of 10 wooden blocks, one green and one red. All blocks were 39 mm high and 39 mm deep with lengths 19 mm, 32 mm, 45 mm, 58 mm, 71 mm, 84 mm, 97 mm, 110 mm, 123 mm, 136 mm. A heavy cloth was used to cover stimulus arrays set up prior to the beginning of a test session.

Design

In the first session subjects received a ten-item classical seriation pre-test. The criteria for success were that the subject (1) reproduce the model with not more than two "swapped" blocks and (2) that these must not be more than two ordinal steps apart. "Reflected" productions, in so far as they meet the above criteria, were counted as correct. If subjects did not meet these criteria, then the number of blocks was reduced until a number which could be successfully seriated was found. For sets less than ten, perfect monotonic seriations were the criterion for success. If subjects could not seriate at least five blocks, not further reduction was made and the subject exited the study.

Procedure

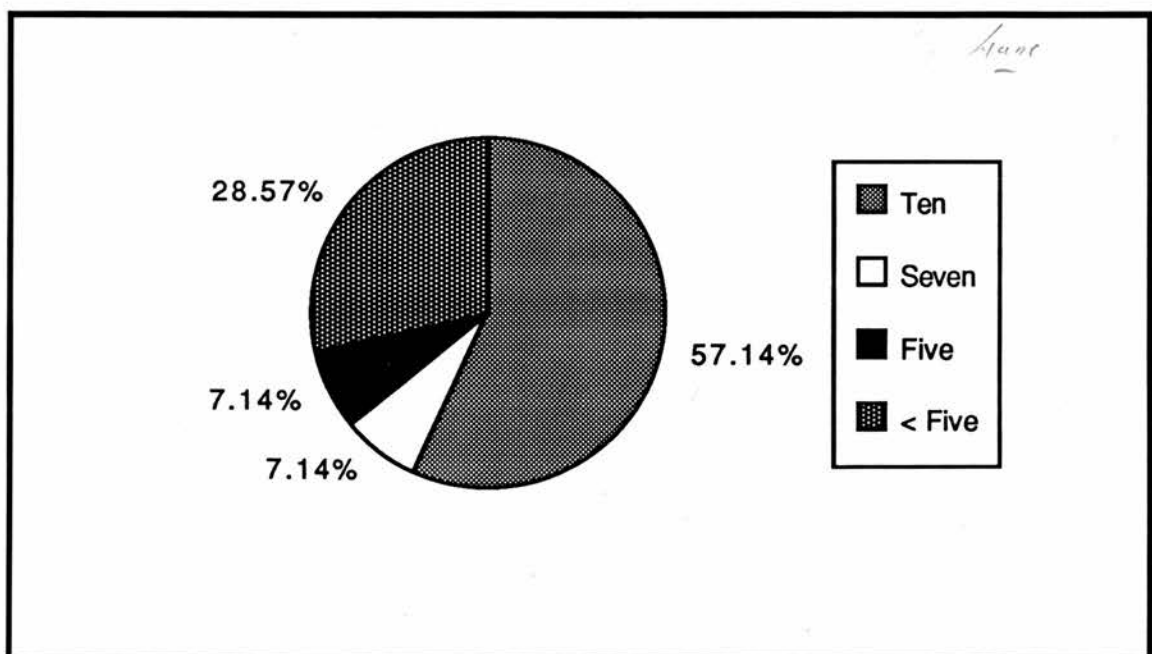
Ten-Item Seriation Pretest: When the subject entered the room a seriated set of blocks was on the table covered by a cloth. Subjects were invited to sit at the table and the experimenter sat down opposite the subject. The experimenter introduced the task as follows: "Today's game goes like this. I have some blocks underneath this cover and when I take it off let's see if you can make one just like mine here with some blocks that I will give you." The cover was removed and subjects were given a set of blocks

which matched the model. Subjects' blocks were arranged in an unordered pool (no block was on top of any other block) away from the area where they were to place them for seriation. The blocks were placed along a ruler which was attached to the table.

When subjects indicated that they were finished, the experimenter asked subjects if their arrangements were "just like" the model. Subjects' responses were noted. If subjects reported that their models were not just like the experimenter's, subjects were given an opportunity to repair the production. However, no specific intervention was given. If subjects met the success criteria, they were congratulated on their performance and left the study here. If subjects did not meet the success criteria, they immediately proceeded to the reduced set task below.

Seriation with Reduced Sets: The above procedure was repeated with five blocks (using the same set of blocks, but omitting the five smaller blocks). In the case of seriation with sets of less than ten items, the success criterion was a perfect monotonic seriation. If subjects did not meet this criterion with five items, then they exited the study here. If subjects met the success criteria with five items, then the procedure was repeated with seven blocks, then nine blocks, etc. until subjects failed to meet the criterion.

Results



Breakdown of performance

The chart above shows the breakdown of performance in terms of maximum set size for the group. On the seriation pretest, whilst 57% of subjects succeeded with ten items, 43% failed. Of these subjects, two-thirds also fail to succeed at seriation with as few as five blocks. There were only two subjects who successfully and reliably seriated a reduced set (five and seven blocks) after a failure at ten. Neither of these subjects showed a reliable, principled executive for controlling their actions.

Appendix C: Limited Review of Monotonic and Non-monotonic Series

EXPERIMENT

Subjects

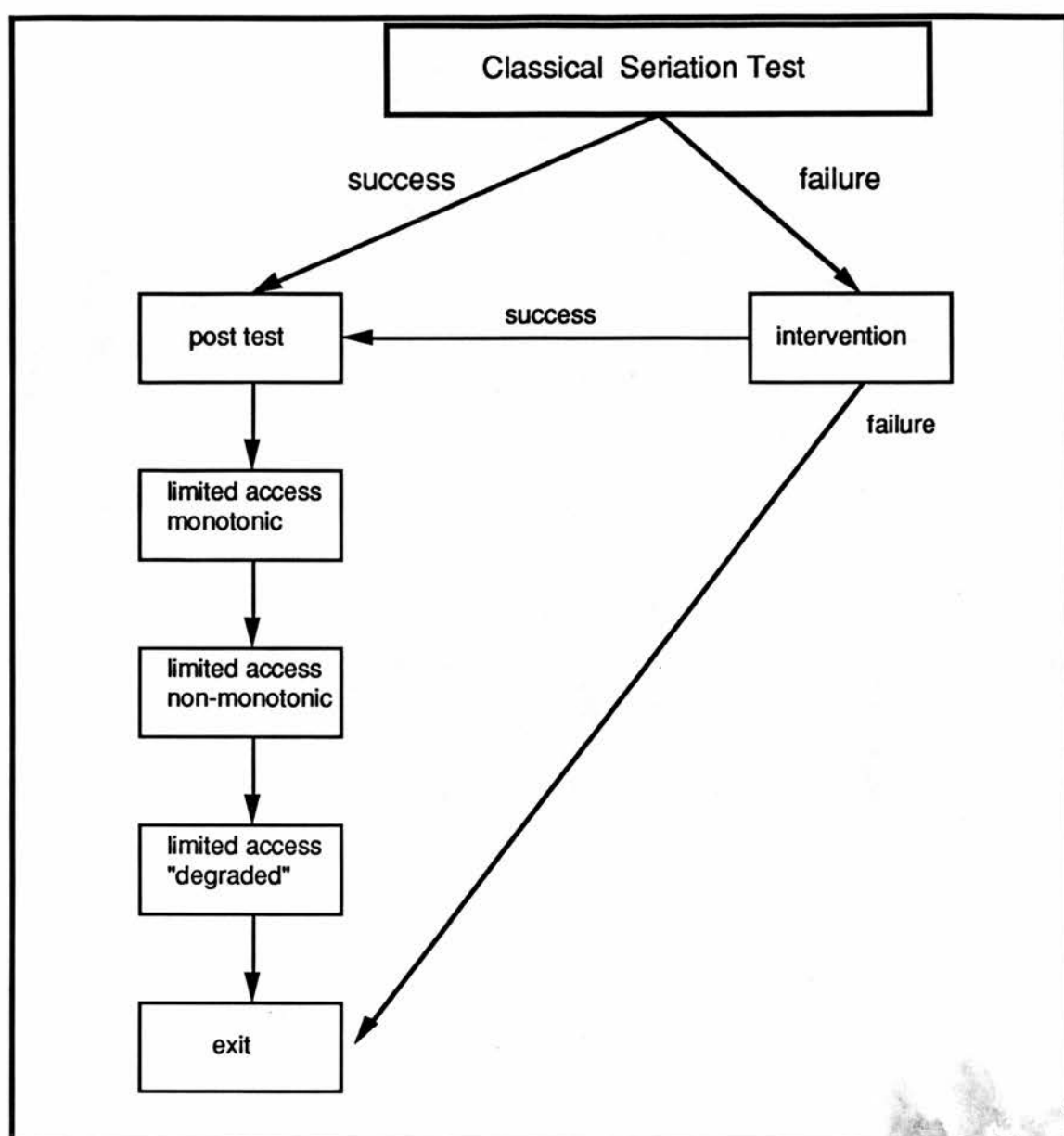
8 four year olds at Edinburgh University Psychology Department Nursery, 11 five year old at Bruntsfield Primary School, and 20 six year old children at Sciennes Primary School.

Materials

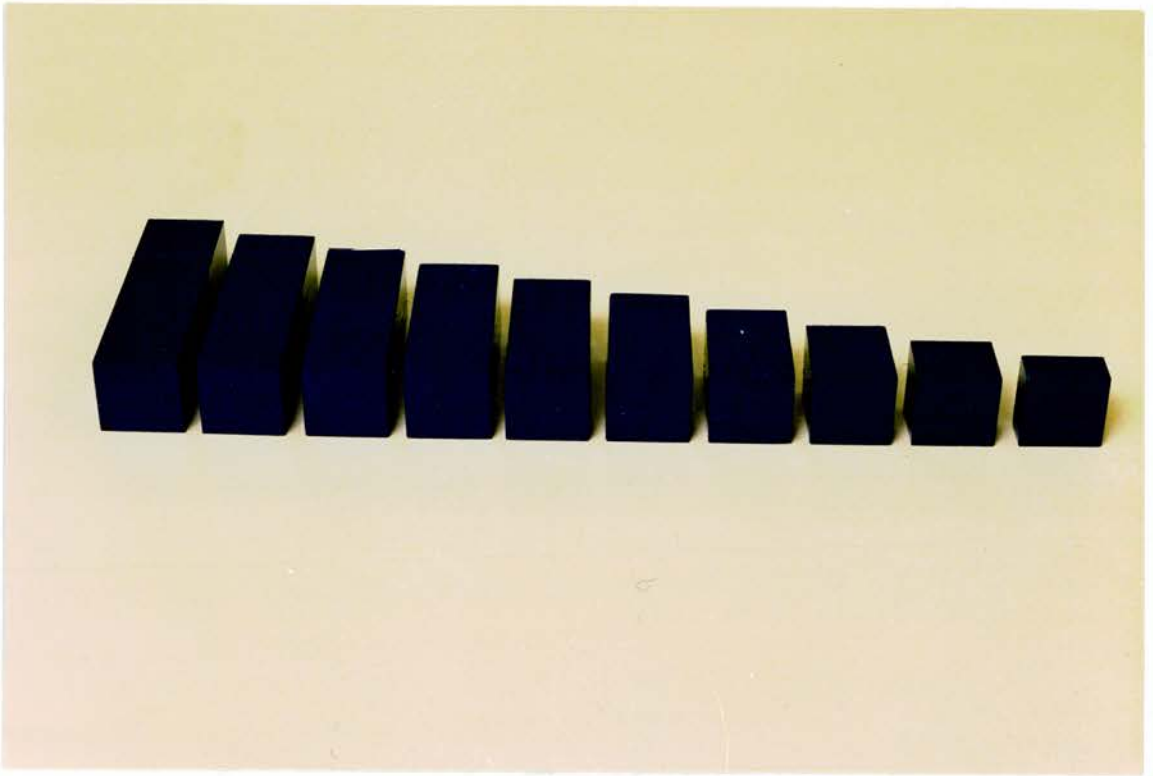
2 sets of 10 wooden blocks, one green and one red. All blocks were 39 mm high and 39 mm deep with lengths 19 mm, 32 mm, 45 mm, 58 mm, 71 mm, 84 mm, 97 mm, 110 mm, 123 mm, 136 mm. A screen was used to occlude the experimenter's actions from the subject when changes in the stimulus array were required during a test session. A heavy cloth was used to cover stimulus arrays set up prior to the beginning of a test session.

Design

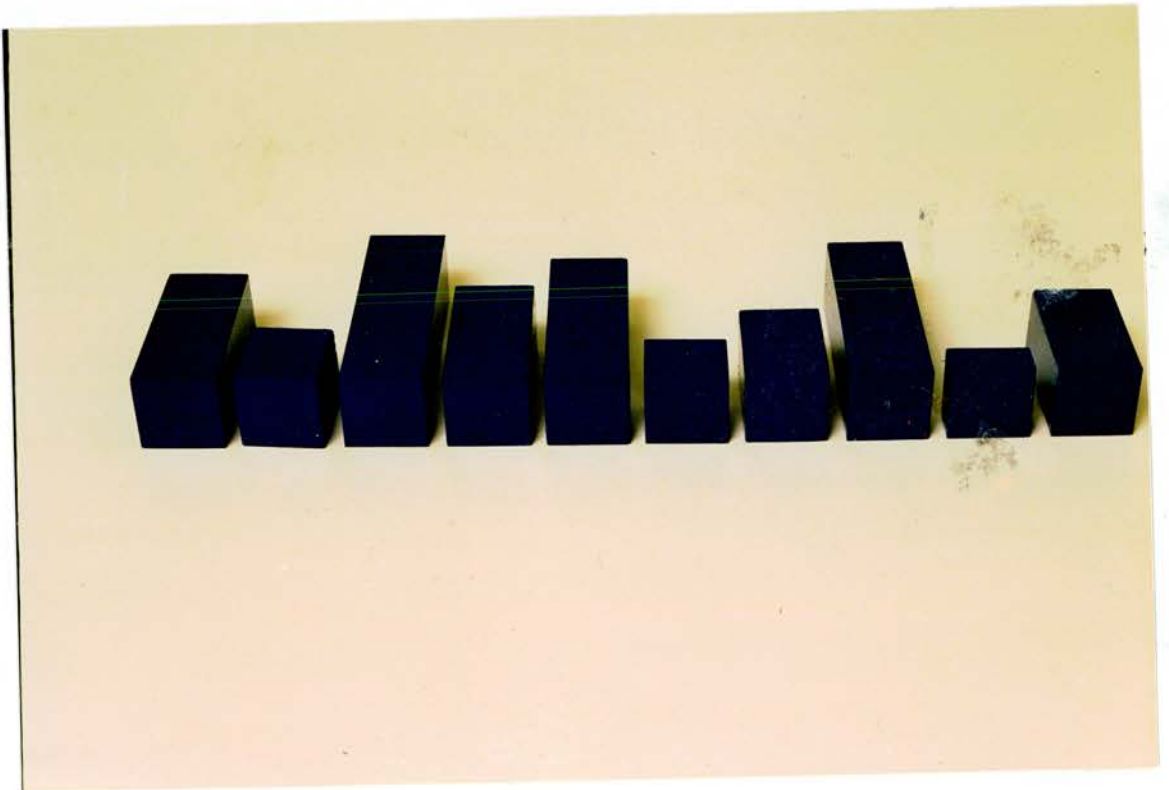
Each subject received the following tasks in the order specified below. Tasks were not counterbalanced because of the questions being addressed. If subjects were given the non-monotonic series before the monotonic series, they might be cued into using a copying strategy to construct the monotonic series where they otherwise might use a principled production.



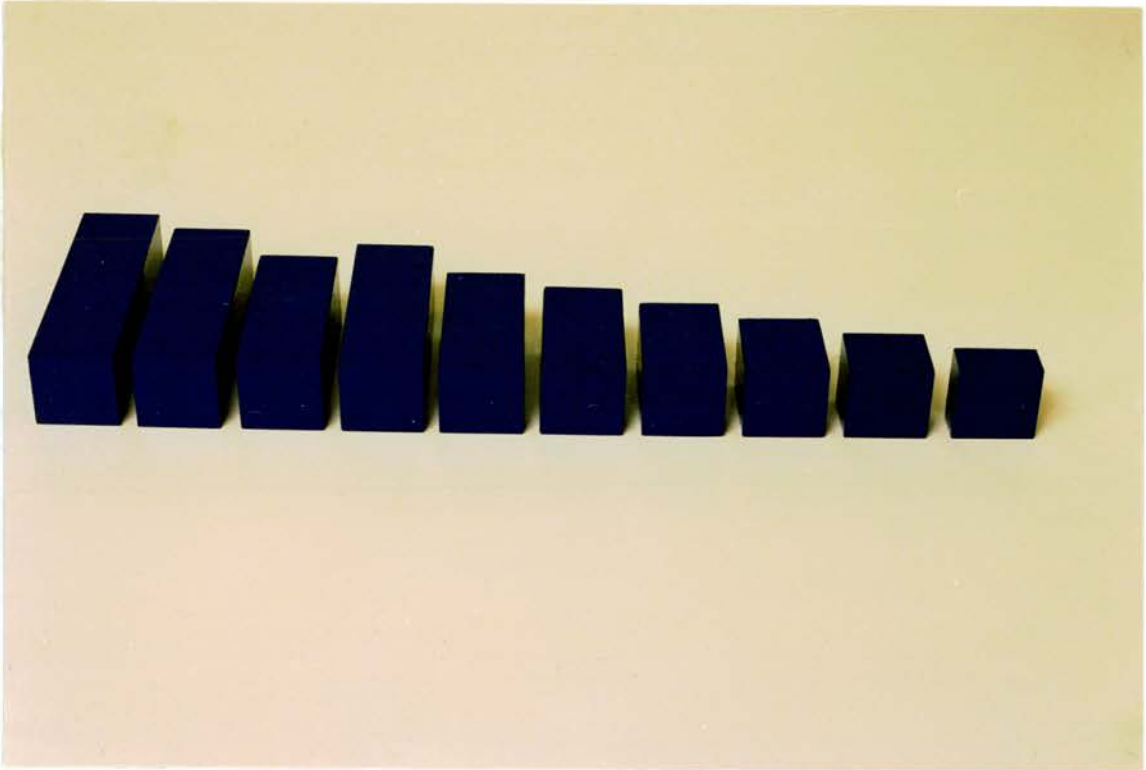
Design



The Monotonic Series



The Non-Monotonic Series



The "Degraded" Series

Procedures

Seriation Pre-test: The basic seriation task, intervention procedure, and seriation retest were administered as per the extended clinical investigation (see Chapter 2). These tasks were administered in one session of approximately 20 minutes.

Limited Model Access: The monotonic and non-monotonic tasks were administered in one session. The "degraded" task was administered in a separate session. The non-monotonic model was always constructed to the specification 4, 8, 1, 5, 3, 9, 6, 2, 10, 7 (see photograph above). The degraded model was always constructed to the specification 1, 2, 4, 3, 5, 6, 7, 8, 9, 10. This particular minimal degradation was chosen because it was found that systematic seriators often made errors of this kind, especially at ordinal positions 3 and 4.

When the subject entered the room the seriated model was on the table covered by a cloth. The task was introduced as follows: "We're going to

play a different game this time. I've got a secret one under here now. Let's see if you can make one just like mine with your blocks, but this time the game will be different. You can look at it for as long as you like, and you tell me after you have seen it long enough and then we'll cover it up and you can make one just like it with the blocks I give you. If you get stuck and can't remember, you can ask to look at it again for as long as you like, and as many times as you like, until you make yours just like mine. But when you are looking at the model you mustn't touch the blocks because that's cheating." The cloth was removed and when subjects indicated that they had "seen it enough," the experimenter re-covered the series. She then placed the subject's blocks on the table in an unordered pool.

Upon completion of their productions, subjects were asked to review: "Is yours just like mine?" Subjects received no differential feedback and immediately proceeded to the next task.

When it was necessary to construct a new model when the subject was in the testing room, the experimenter screened the model area from the subject's view. The "rules" were repeated between tasks to refresh subjects' memories

Results

AGE (NUM)	MONOTONIC	NON-MONO	DEGRADED
4 (08)	100%	0%	0%
5 (11)	91%	0%	0%
6 (20)	90%	0%	30%

Percent correct for each series

There was a striking difference between subjects' abilities to construct a monotonic model versus non-monotonic models. There was universal failure with the non-monotonic model regardless of age. Subjects were also largely unsuccessful in constructing the degraded series, with only 30% of six year olds being able to correctly produce it.

AGE (NUM)	CORRECT	MONO FAIL	OTHER FAIL
4 (08)	0%	75%	25%
5 (11)	0%	36%	64%
6 (20)	30%	35%	35%

Distribution of performance types on the degraded monotonic

Although few subjects were able to accurately reproduce the degraded model, many subjects erred in favour of redundancy, producing a perfect monotonic model rather the degraded model presented to them.

All subjects verbally expressed a preference for the monotonic series, which was considered to be "easier." Younger subjects typically expressed a reluctance to attempt the non-monotonic series (one subject refused to do it at all).

AGE (NUM)	MONOTONIC	NON-MONO	DEGRADED
4 (08)	1.13	3.00	1.25
5 (11)	0.27	6.45	2.09
6 (20)	0.00	4.15	1.85

Average number of requested reviews per series

Subjects' requests for additional access to the model was related to model type. In the case of the monotonic series, most subjects were able to complete the model correctly after the initial viewing. Only two subjects, one four year old and one five year old, requested to review the monotonic model. In the case of the non-monotonic model, where there was no success at any age, subjects request many additional viewings. There were fewer requests to review for the degraded model, however this was largely due to subjects' tendency to error in favour of the monotonic. For many subjects it was not clear that they actually saw a relevant difference between the monotonic series and the degraded series. One subject, after producing a monotonic series and assuring the experimenter that it was correct, commented that he thought the model "wasn't quite right."

There were some age-related differences in requests to review, especially with respect to how subjects used their additional reviews of the model. The youngest subjects seemed to be unwilling to participate in the whole

review and correction process and were content with their productions after one viewing. With respect to *use* of additional reviews, younger subjects (aged four and five years) typically put out all ten blocks after the initial inspection of the model and then used their additional reviews to modify the production as a whole. Six year olds, on the other hand, tended to "chunk" their productions, requesting further inspections of the model before they had laid out the entire complement of ten blocks. On average, six year old subjects managed four blocks per inspection.

These results suggest that successful seriators (with respect to a ten item set) do not construct a monotonic series by copying it "on-line". Rather, they produce it from memory, and are likely to be taking advantage of the redundancy inherent in the model. Once these redundant features are taken away (as in the non-monotonic model), subject must rely on a more difficult means of producing the series, direct copying. Older subjects become more sophisticated at this strategy when they are required to use it; they break up the production into more manageable "chunks." Direct copying seems to be very difficult for young children.